Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *Human Resources Research Center*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

LIBRARY
SEP 4 - 1953

CONTENTS

The Effect of Speed and Load on Display-Control Relationships: W. B. KNOWLES, W. D. GARVEY, AND E. P. NEWLIN 65
Some Characteristics of the "Range Effect": C. W. SLACK 76
Psychological Probability as a Function of Experienced Frequency: F. ATTNEAVE 81
Effect of Amount of Verbal Anchoring and Number of Rating-Scale Categories upon Transmitted Information: A. W. BENDIG AND J. B. HUGHES II 87
Two-Channel Listening: E. C. POULTON 91
A Retest for Conditioned Inhibition in the Alphabet-Printing Task: R. E. SCHUCKER, L. B. STEVENS, AND D. S. ELLIS 97
Patterning Effect in Partial Reinforcement: J. H. GROSSLIGHT, J. F. HALL, AND J. MURNIN 103
The Development of Noncontinuity Behavior through Continuity Learning: L. S. REID ... 107
Magnitude of Reward and Acquisition of a Black-White Discrimination Habit: J. E. GREENE 113
Anxiety and Task as Determiners of Verbal Performance: C. K. RAMOND 120
A Psychophysical and Electrophysiological Study of Light Adaptation: R. M. BOYNTON AND M. H. TRIEDMAN 125
Comment on "The Role of Instruction in Experimental Space Orientation": H. A. WITKIN 135

American Psychological Association

Vol. 46 No. 2 August 1953

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors who request them are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas. Priority in publication is given to articles whose authors assume the full cost of publication.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 46, NO. 2 AUGUST, 1953

THE EFFECT OF SPEED AND LOAD ON DISPLAY–CONTROL RELATIONSHIPS [1]

WILLIAM B. KNOWLES, WILLIAM D. GARVEY, AND EDWARD P. NEWLIN [2]
Naval Research Laboratory

The experiments reported here are part of a series of studies aimed at discovering better ways of designing display-control systems as efficient functional units. The general problem has been considered from two points of view. First, several aspects of display-control interaction have been studied. Second, the experiments were designed so the results could be analyzed within the framework of communication theory.

In relation to the first approach, Fitts [3] and his co-workers (4) have recently reported evidence to show that the efficiency of a given control may be determined to a large extent by the display with which it is combined and vice versa. For his displays, Fitts used lights arranged in three simple patterns. Paired with each of the displays were three stylus-and-slot controls. In each of the resulting nine display-control combinations it was possible to designate, with the lights, eight different positions to which the styli were to be moved. The analysis of variance of the response times showed that the *F* ratios for displays and for controls were not significant, but that the *F* ratio for controls-by-displays interaction was highly significant. Fitts used the term "compatibility" to express the idea that displays and controls may interact with one another to produce more efficient systems than would be expected from a knowledge of the efficiencies of the individual displays and controls. It seemed desirable to investigate display-control interactions and how they come about in somewhat more complex systems than those used by Fitts.

The second, or communication-theory approach, was taken because it was thought that new insights into the problems of designing display-control systems might be obtained by considering the display-man-control system as a communication channel, and, then, by attempting to describe the characteristics of this channel. The emphasis here is upon stimulus and response probabilities as important variables influencing the oper-

[1] The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

[2] The writers are indebted to Dr. Franklin V. Taylor for aid in formulating the problem, and to Mr. Alan W. Baldwin for assistance in the construction of the apparatus.

[3] Fitts, P. M. Personal communication. Sept., 1951.

ator's performance. These probabilities are translated, in communication-theory terms, into amounts and rates of transmitted information.

The experiments presented here were designed to answer the following questions: (*a*) Can compatibility be demonstrated between displays and controls which approach the level of complexity found in actual operating situations? (*b*) How does changing the size (load, or the number of possible states) of the systems affect their efficiencies? (*c*) How accurately can the systems be operated when signals are presented to the operator at increasingly faster rates?

Method

Apparatus.—The displays and controls used in these experiments are shown in Fig. 1. The displays were (*a*) a matrix of lights, (*b*) a window, or modified memory-drum-type display, and (*c*) spoken words presented over a loud-speaker from magnetic tape recordings. The controls were (*a*) a matrix of push buttons, (*b*) a row and column of push buttons, and (*c*) a double column of push buttons. The controls were under consideration for use in a specific operating situation. The displays were selected to represent prototypes of displays which might be used in the same kind of operating situation. The control panels and matrix display were painted black except for the white reference indexes around the margins. Orange plastic buttons spaced 1 in. apart from center to center were used on the control panels and neon lights, similarly spaced, were used on the matrix displays.

The studies simulated situations in which the operator is required to connect one set of elements, the letters, with another set of elements, the numbers. For example, if the light in Row A, Column 1 on the matrix display came on, this indicated that A was to be connected with 1. The same signal would appear as the printed symbol "A-1" on the window display and as the spoken words "A, one" over the auditory display. These stimuli were responded to by pressing the single button in Row A, Column 1 on the matrix control. On either the row and column or double-column control, the response would be to press the "A" and "1" buttons simultaneously.

The display-control systems were set up in a sound-shielded experimental room remote from *E* and the main control and recording apparatus. The *S*s were seated at a small table on which the control panels were placed. The table was adjusted to a comfortable height and the top was tilted to about 35° from the horizontal. The displays were centered above the controls on a small shelf at approximately eye level and about 30 in. from *S*. The walls and ceiling of the room were white; fluorescent ceiling lights provided a moderately bright general illumination.

A number of predetermined random sequences of stimuli were presented by means of a specially constructed, punched tape, programming device. This apparatus also cumulated response times and errors over the trial sequences. Polygraph records were taken so that response times and errors for the individual stimuli could be measured. The programmer allowed for two modes of stimulus presentation: (*a*) self-pacing, where each response immediately caused the next stimulus to appear, and (*b*) forced pacing, where an adjustable interval timer automatically presented the stimuli at given rates.

*Experiment 1*a.—The purpose of the first part of this experiment was simply to demonstrate the possible interactions between the three controls and three displays shown in Fig. 1. Five Navy enlisted men stationed at the Naval Research Laboratory and four members of the psychology branch staff served as *S*s. Before the first experimental session, *S*s were told that the purpose of the experiment was to evaluate the various display-control systems. They were then instructed on the operation of all the combinations and were given one practice trial of 50 stimuli on each of the controls using the matrix display. They were further instructed to work as quickly and as accurately as possible and not to attempt to correct errors. The self-pacing mode of stimulus presentation was used.

Three displays paired with each of three controls gave nine display-control conditions. Each *S* was given one experimental trial of 100 stimuli on each of the conditions according to a randomized 9×9 latin-square design. There were 100 possible stimuli and these were presented in three different random sequences. Cumulative response time and errors per 100 stimuli were the measures taken.

*Experiment 1*b.—The purpose of this part of the experiment was to determine the effect of reducing the size, or number of stimulus alternatives, on the efficiencies of four of the original display-control combinations. The matrix and window displays in combination with the matrix and the row and column controls were the conditions selected on the basis of the results of the first part of the study. The number of stimulus alternatives was reduced from 100, as in Exp. 1*a*,

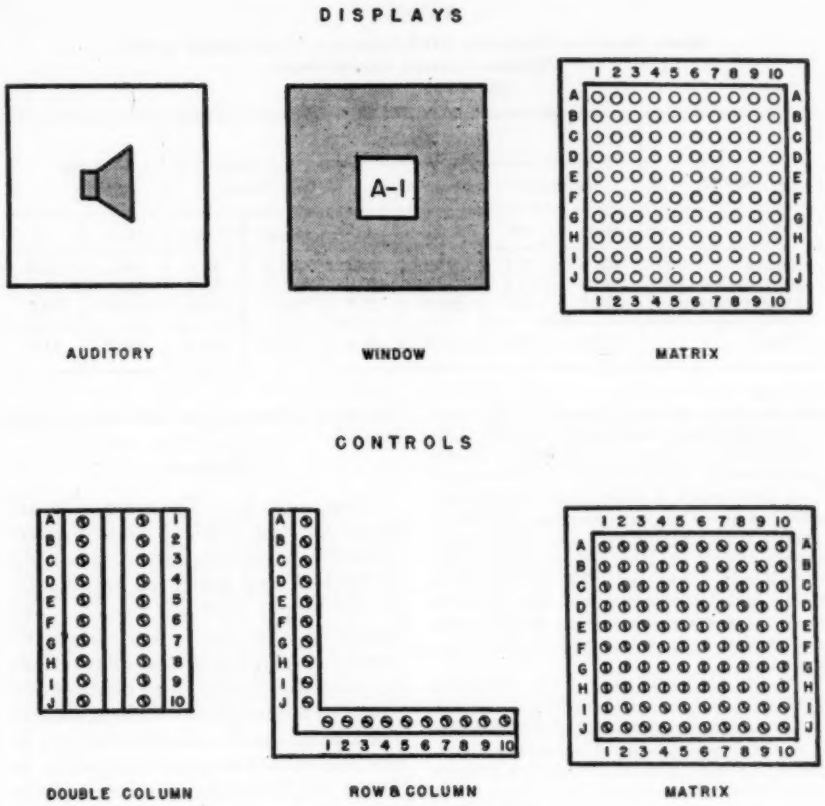


FIG. 1. Displays and controls

to 25 by masking off all but the lower left quadrant of the matrices and the corresponding buttons on the row and column control. New memory-drum tapes were prepared for the window display.

The *S*s were a different group of eight Navy enlisted men stationed at the laboratory. They were given the same instructions and practice as in Exp. 1*a*. Each *S* was then given one experimental trial on each of the four conditions. This set of four trials was followed by a second set of trials in which the conditions were presented in the same order. The orders of presenting the conditions were determined from a 4×4 counterbalanced latin square. All told, this basic square was replicated once with repeated measurements taken on both replications. The purpose of this design was to control for the general practice effect found in the first experiment while using essentially the same number of *S*s. The self-pacing mode of stimulus presentation was used with *S*s operating under speed and accuracy instructions. Cumulative response times and errors per 100 stimuli were the measures taken.

Experiment 2.—The purpose of this experiment was to determine how accurately the display-control combinations of Exp. 1*b* could be operated under two conditions of load combined with four conditions of speed. The matrix-matrix, matrix-row and column, window-matrix, and window-row and column combinations were used in the 100- and 25-stimulus alternative sizes. Stimuli were presented at rates of .27, .34, .43, and .45 stimuli per second on the large forms, and of .38, .49, .60, and .65 stimuli per second on the small forms. These rates of

TABLE 1

MEAN RESPONSE TIMES PER 100 STIMULI FOR LARGE FORMS OF THE DISPLAY–CONTROL COMBINATIONS

($N = 9$ each condition)

Controls	Displays						Total	
	Auditory		Window		Matrix			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Matrix	166.8	31.3	223.4	23.4	207.9	26.3	199.4	36.0
Row and column	155.0	22.5	217.9	20.9	270.0	32.2	214.3	53.6
Double column	159.5	19.7	209.8	25.7	287.7	34.2	219.0	59.4
Total	160.4	25.6	217.0	24.4	255.2	46.3	210.9	51.3

stimulus presentation corresponded to the rates of informational presentation of 1.75, 2.25, 2.75, and 3.00 bits per second on both forms.[4] Errors per 100 stimuli was the measure of performance.

A third group of eight Navy enlisted men served as *S*s. They were given one preliminary practice trial of 50 stimuli on each of the combinations on both the large and small sizes under the self-pacing conditions. The speeds selected for use in the experimental trials were chosen to bracket the average self-pacing speed on the slowest combination calculated from these practice trials.

Four display-control combinations in two sizes yielded eight experimental conditions. Each *S* was given one session on each of the conditions in an 8 × 8 counterbalanced latin-square design. During a session, four trials, each at a different rate of stimulus presentation, were given. The orders of presenting the rates were determined from a counterbalanced 4 × 4 latin square. A single trial, then, consisted of a random series of 100 stimuli presented on a given display-control combination of a given size at a given rate. The measure of performance was the number of errors per trial.

RESULTS

*Experiment 1*a.—The mean response times for the original nine display-control combinations are summarized in Table 1 and presented graphically in the upper part of Fig. 2. The points in this figure are connected merely to emphasize the interactions between the displays and controls; no continua are implied. This figure and the significant *F* ratio for displays in the analysis-of-variance summary in Table 2 indicate that there were large differences among the various combinations attributable to the displays. In general, the auditory display was

TABLE 2

ANALYSIS OF VARIANCE OF RESPONSE TIMES FOR THE LARGE FORMS

Source of Variance	*df*	Variance Estimate	*F*
Trials	8	1,145.12	10.36**
*S*s	8	4,738.32	44.44**
Displays	2	61,417.87	575.99**
Controls	2	2,843.55	26.67**
D × C	4	6,843.85	64.18**
S × D	16	215.45	2.02
S × C	16	292.64	2.74*
Error	24	106.63	

** $p < .01$.
* $p < .05$.

[4] Assuming that, from *S*'s point of view, the probabilities of occurrence for each stimulus are equal and independent, the amount of information per stimulus, H, in bits per stimulus, is given by: $H = \log_2 n$, where n equals the number of possible stimuli. Therefore, when $n = 100$, $H = 6.6$ bits per stimulus and when $n = 25$, $H = 4.6$ bits per stimulus. The average rate of informational presentation, H', in bits per second, is given by $H' = mH$, where m equals the average rate of stimulus presentation (5, p. 23). The rates of stimulus presentation given here are the actual calibrated values used in the study. The rates of informational presentation given approximate the calculated values and are accurate only to the first decimal place.

TABLE 3

MEAN RESPONSE TIMES PER 100 STIMULI FOR SMALL FORMS OF THE DISPLAY–CONTROL COMBINATIONS

(N = 16 each condition)

Controls	Displays				Total	
	Window		Matrix			
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Matrix	135.5	28.9	110.4	19.1	123.0	27.5
Row and column	154.6	25.6	171.7	25.2	163.1	26.8
Total	145.0	28.8	141.1	37.9	143.0	33.8

better than either of the two visual displays. Of the two visual displays, the window, on the average, permitted more efficient operation than the matrix.

The differences attributable to the controls are not quite so clear-cut. From the significant F ratio for controls and a comparison of the control means, it might be concluded that the matrix control was better than either the row and column or the double-column controls. However, this conclusion must be considerably qualified in light of the significant displays-by-controls interaction. A detailed analysis of the differences between the controls when combined with the various displays in terms of t tests shows the following: (*a*) When combined with the matrix display, the matrix control was better than either of the other controls ($p < .01$); (*b*) when combined with the window display, the three controls were equivalent; (*c*) when combined with the auditory display, the matrix control was poorer than the row and column control ($p < .05$), but did not differ significantly from the double-column control; (*d*) in all cases, the row and column and double-column controls were equivalent.

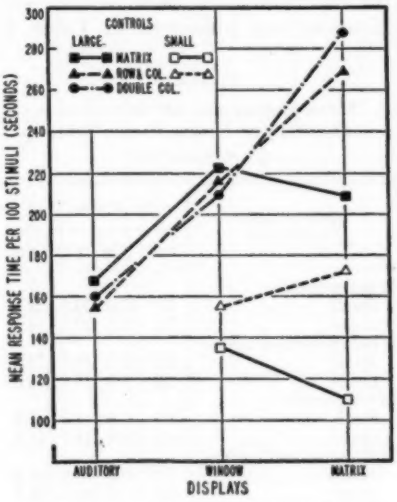


FIG. 2. Display-control interaction

Errors proved not to be significantly different from zero and, therefore, are not reported.

These results confirm Fitts's finding that the efficiency of a given control or display may be determined to a large extent by the display or control with which it is associated. It has been shown that the matrix control, which was equivalent to or poorer than the other controls when used with either the window or auditory displays, became by far the best control for use with the matrix display. In Fitts's terminology the matrix-

TABLE 4

ANALYSIS OF VARIANCE OF RESPONSE TIMES FOR THE SMALL FORMS

Source of Variance	*df*	Variance Estimate	*F*
Trials	7	364.45	2.72*
Ss	7	4,128.87	30.82**
Displays	1	246.88	1.84
Controls	1	25,828.51	192.82**
D × C	1	7,129.69	53.23**
S × D	7	261.08	1.95
S × C	7	312.58	2.33*
Error	32	133.95	

* $p < .05$. ** $p < .01$.

matrix combination is a "compatible" system.

*Experiment 1*b.—The results obtained by reducing the size of four of the systems are presented in Table 3 and are also plotted in the lower part of Fig. 2. The differences between the repeated measurements on the conditions were not significantly different from zero as shown by matched *t* tests. Therefore, the data for all observations are combined and each point represents the mean of 16 readings. The analysis-of-variance summary in Table 4 shows that there was no difference between the small matrix display and the window display. The small matrix control was significantly better for use with either display than was the small row and column control. However, displays-by-controls interaction was also significant. Analysis by *t* tests shows that the matrix control combined with the matrix display was better than the matrix control combined with the window display ($p < .01$), but that the row and column control was poorer when combined with the matrix display than when combined with the window display ($p < .01$).

In general, the relative efficiencies of the systems are quite similar for both the large and small forms and altering the size of the systems does not appear to alter the display-control interrelationships appreciably, at least over the range used in these studies.

Changing the size of the systems does change the rates at which stimuli can be handled. In a reaction-time study similar to these, Hick (**3**) has shown that the rate at which *stimuli* are handled is inversely related to the number of stimulus alternatives, but that the rate at which *information* is transmitted remains constant—at least over the range of from 1 to 10 stimulus alternatives. The rate of transmission of information, in bits per second, was calculated for each form of the four display-control combinations (see Table 6). It was found, using the *t* test for independent groups, that significantly less information was transmitted per second on the large form of the matrix-matrix combination than on the small form ($p < .01$). The differences between the rates of transmission on the two forms of each of the other combinations were not significant.

Experiment 2.—The data showing the effects of speed stress on the combinations are presented in Table 5.

TABLE 5

MEAN ERRORS PER 100 STIMULI AS A FUNCTION OF FORCED-PACING SPEED ($N = 8$ each cell)

Display-Control Combination	Speed (Bits/Sec.)			
	1.75	2.25	2.75	3.00
Window-matrix				
Small	5.8	13.9	23.5	48.8
Large	5.1	9.2	27.1	35.9
Window-row and column				
Small	2.9	20.0	39.8	71.8
Large	2.0	3.8	25.8	38.0
Matrix-matrix				
Small	2.5	3.6	4.1	7.1
Large	3.6	10.8	13.1	15.8
Matrix-row and column				
Small	5.2	30.0	47.9	61.6
Large	6.6	23.8	44.5	52.1

Statistical sign tests (1) of the differences in errors between the various large systems taken over all speeds show that (*a*) more errors were made on the large matrix-row and column system than on any of the other large systems ($p < .01$), and (*b*) the other large systems were not significantly different from one another. A similar comparison of the small forms shows (*a*) significant differences between all the combinations ($p < .05$), except that (*b*) the small window-row and column and matrix-row and column combinations were not significantly different. These results indicate that the same general relationships are found among the systems when performance is measured in terms of errors as when response time is used as the measure. They show, furthermore, that speed stress affects the incompatible systems more than the compatible systems.

The error vs. speed curves for the compatible matrix-matrix and the incompatible matrix-row and column combinations have been plotted in Fig. 3 with speed being measured in terms of *stimuli per second*. The curves for the other combinations are omitted to simplify the figure but they would fall, generally, between the extremes presented. Inspection of the curves in Fig. 3 shows that at any specified rate of stimulus presentation fewer errors were made on the smaller forms of the combinations. The differences in the error trends of the various large combinations were tested using Friedman's (2) analysis-of-variance technique with ranked data. These tests show that (*a*) accuracy degenerates less with increasing speed on the matrix-matrix system than on any of the other large systems ($p < .01$), and that (*b*) the trends for the other large systems are essentially the same. The same type of trend analysis for the small forms shows that (*a*) accuracy degenerates less on the matrix-matrix combination than on any of the other small systems ($p < .01$), and (*b*) the trend for the small window-matrix system is significantly

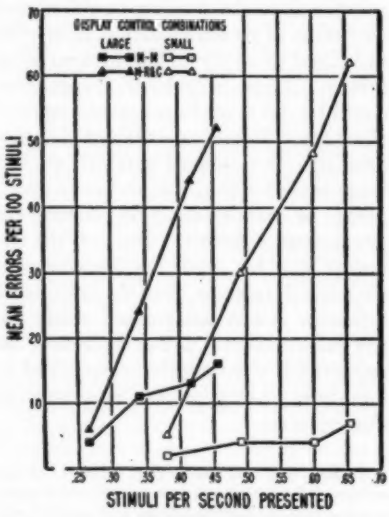


FIG. 3. Error as a function of rate of stimulus presentation

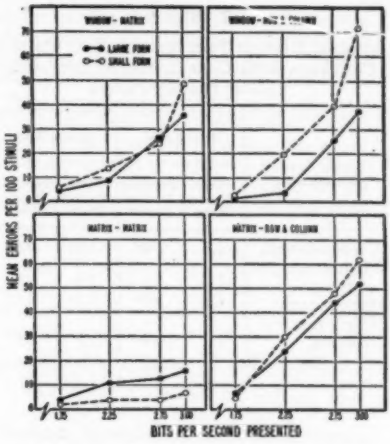


FIG. 4. Error as a function of rate of informational presentation

TABLE 6

AVERAGE RATE OF TRANSMISSION, SELF-PACING, COMPARED WITH CORRESPONDING ESTIMATED ERROR, FORCED-PACING

Display-Control Combination	Small Size		Large Size	
	Self-Pacing Rate	Forced-Pacing Error	Self-Pacing Rate	Forced-Pacing Error
Window-matrix	3.5	70*	3.0	36
Window-row and column	3.0	72	3.0	38
Matrix-matrix	4.3	10*	3.2	18*
Matrix-row and column	2.7	44	2.5	33

* Obtained by extrapolation of curves in Fig. 4.

different from that of the small matrix-row and column combination ($p < .05$).

In Fig. 4 the large and small forms of all the combinations are compared directly by equating the speed stress in terms of rate of *informational* input, i.e., bits per second. Statistical sign tests were performed to test the differences between the error scores of the large and small forms of each system over the entire range of speeds presented. Significantly fewer errors were made on the smaller matrix-matrix system ($p < .01$), while significantly more errors were made on the smaller form of the window-row and column system ($p < .01$). Friedman's analysis-of-variance tests for the differences in trends also indicated significant differences in the error functions for these two systems ($p < .05$).

A comparison of the self-pacing data with the forced-pacing data indicates that more information is transmitted per unit time in the self-pacing mode of operation. These data are shown in Table 6. This table was constructed by calculating the rate at which information was handled in the self-pacing situations and then estimating from the curves in Fig. 4 the number of errors to be expected had *Ss* been forced to respond to information presented at the calculated rates. Although three of the forced-pacing entries were obtained by extrapolating the empirical curves, this only serves to emphasize the superiority of the self-pacing mode of operation. The data in Table 6 also show that speed stress does not degrade performance as much on the compatible combination as on the other systems.

DISCUSSION

Before discussing these results in relation to compatibility and their implications for design practice, it will be desirable to attempt to answer the question: Can the operator's behavior in these situations be adequately described in terms of the properties of a communication system as developed by Shannon (5). The consideration of this question is important not only from a theoretical point of view but for the establishment of a frame of reference for the subsequent discussion of compatibility. To answer this question most satisfactorily would require that the basic postulates and theorems of communication theory be redefined to apply to the present situations. Then a set of hypotheses could be developed and tested against the data. However, it is not the purpose of this paper to attempt such a large-scale redefinition and redevelopment of communication theory in psychological terms. Rather, the attempt will be made to demonstrate that the data show that the operator's behavior is not determined solely by the informational input as would be expected if the operator responded in a manner analogous to a communication system.

In Exp. 1 it was shown that for three of the systems information was transmitted at the same average rate on both the large and small forms when *Ss* were self-paced. However, in the case of the matrix-matrix combination more infor-

mation was transmitted per second on the small form than on the large. Communication theory uses the term "information" somewhat synonymously with "choice," "uncertainty," and "entropy." It is helpful in interpreting the above results to think of information as a measure of the amount of choice involved in responding to a signal as Hick (3) has done. Hick's studies show, essentially, that as the number of alternatives from which *S* must choose increases, the average response time for each alternative increases. But, he shows further that the response times increase as a logarithmic function of the number of alternatives, or in such a way that the rate of transmission of information remains constant. In other words, in Hick's studies, his *S*s "chose" or "reduced uncertainty" at a constant rate regardless of the number of possible alternatives. This would seem to indicate that the basis for choosing, or the kinds of choices, remained the same or were equivalent under all conditions, and that the greater time required for the selection of one out of many alternatives reflected the greater number of units of choice involved in making the selection. Another way of putting it is to say that the method of "coding" the information remained the same or was equivalent throughout the range of alternatives.

The equal rates of transmission of information found for three of the systems in Exp. 1 are interpreted in the same way, i.e., the same or equivalent coding methods were used on both forms. The faster rate of transmission for the small matrix-matrix system implies that a different and easier coding method was utilized on this form than on the larger form. It also implies that *S* changed his characteristics as a communication channel.

The results of Exp. 2 can also be interpreted in a similar manner. Experiment 2 was designed so that information could be presented to *S* at four different rates, 1.75, 2.25, 2.75, and 3.00 bits per second. By using the large and small forms of the display-control combinations, it was possible to attain each given rate by two methods, e.g., the 1.75 rate was obtained by presenting .38 stimuli per second on the small forms and by presenting .27 stimuli per second on the large forms.

If *S* acts as the same communication channel, and if he employs the same coding method when using the two forms of the same combination, then the same amount of information should be transmitted for equal rates of input on the two forms. No record was kept of the exact response to each stimulus so that it is not possible to calculate the rates of transmission or the exact equivocation functions. But, since equivocation, or the amount of information lost in transmission, is a function of error, the gross error curves in Fig. 4 do give some indications of whether or not significantly different amounts of information were lost at any given rate of input.

These curves show that with two of the four combinations, the window-matrix and the matrix-row and column, essentially the same amount of information was lost at any given rate of input. These results are in line with the expected results if *S* behaves in a manner analogous to a communication system.

However, for the other two combinations there were significant differences between the error functions. The curves for the matrix-matrix combination show that at any rate of input more information was lost on the large form, while the curves for the window-row and column show that more information was lost on the small form. These results are not what would be expected if *S* behaves as a communication channel, for they imply that the coding has changed, or the channel has changed, or both. In any event, it is to be inferred both from these results and those of Exp. 1 that *S*'s behavior was dependent, in part at least, upon variables other than the rate or amount of informational input, and this is not in line with the defined properties of a communication channel.

Further evidence leading to this same conclusion is given by the data in Table 6. These data show that if *S* is forced to respond to the same average rate of input which he establishes for himself in the self-pacing mode of operation, he fails to transmit as much information. This again shows that the rate of informational input is not the sole determiner of *S*'s behavior in these situations.

The above discussion seems to indicate that an analysis of the methods by which *S* takes information from the display and puts it back into the control would be basic to the attempt to develop a systematic theory of display and control design. The general approach to this type of behavioral analysis will be presented more fully in a future paper. But a discussion of the preliminary inquiry into the basis of the matrix-matrix compatibility may serve to indicate the direction of this approach.

The response times for the individual light stimuli on the large matrix-matrix combination were analyzed in terms of central and peripheral stimuli. The outer three rings of lights were considered peripheral and the inner two rings were considered central. The mean response time for the peripheral stimuli was 1.89 sec. per stimulus while the mean response time for the central stimuli was 3.12 sec. per stimulus. Comparing these peripheral and central response times with the data in Fig. 2 shows that the central response times approximate the average

response times obtained with the row and column and double-column controls combined with the matrix display. Likewise, the mean peripheral response time compares favorably with the times obtained with the auditory systems.

A clue as to the cause of this differential in response times was obtained from interviews with *S*s after the experiment was over. According to their reports the central stimuli were identified by referring to the letter and number indexes around the edge of the display. The correct response button was similarly located by referring to the indexes on the control. But when the stimuli appeared in the peripheral area, it was not necessary to refer to the row and column identifications. The task then became a matter of perceiving spatial relationships. For example, if Light A-1 appeared on the light matrix, it was identified as the light in the upper left corner and the response button in the corresponding position on the matrix control was pressed. The reference indexes were ignored.

Thus compatibility, as demonstrated in this experiment, seems to be due to a display-control relationship which permitted the partial use of a relatively simple spatial discrimination in the case of the matrix-matrix combination. The extremely short peripheral response times averaged in with the longer central response times gave a relatively low mean response time for the matrix-matrix combination. Since the spatial-type discriminations were not possible when the other controls were used with the matrix display, these combinations gave considerably longer mean response times.

These observations have led to the consideration of the hypothesis that compatibility is dependent upon the simplicity with which stimuli can be identified and responses can be executed. Furthermore, at least in the present situations, this simplicity seems to be determined by the similarity of the spatial organization of the display with the spatial organization of the control.

In summary, the data presented here lead to the following conclusions and implications relative to compatibility and design practice. The analysis of the response times for the various combinations in Exp. 1 demonstrated that display-control interaction may be a significant and important variable in determining the efficiency of systems which approach the complexity of practical operating systems. The difference between the average response times for the compatible and incompatible systems was some 25 to 30% of the average response times for the incompatible combination. Such a saving, if it can be obtained with other display-control systems, is not inconsiderable. Likewise, the differential effects of speed upon accuracy for the compatible and incompatible systems demonstrated in Exp. 2 show that where accuracy is an important factor, some careful attention should be paid to the possible interactions between the displays and controls selected for the accomplishment of a given task. In short, these results suggest that at the applied level more attention should be given to the design of display-control systems as functional wholes; the assumption that the most efficient display-control combination can be obtained by associating the "best" display with the "best" control appears not to be valid.

At least passing mention should be made of the findings based on the comparison of the self-pacing and forced-pacing results. This comparison showed that the self-pacing mode of operation permitted more stimuli per second to be handled with fewer errors. For example, in the self-pacing mode *S*s averaged .45 stimuli per second with no errors on the large window-matrix system. When *S*s were forced to respond at this rate on the same system, 36% of the responses were in error. The applied implications of these findings are that, whenever possible, sequential operations should be arranged so that the operator can govern the rate at which the individual steps of the operation are presented.

There is some reason to believe that as display-control systems become more complex, the relative importance of compatibility may diminish. In Exp. 1 it was shown that increasing the size of the systems reduced the efficiency of performance in the sense that more time was required to handle a given number of stimuli, although changing the size did not change the relative efficiencies of the systems. However, when the same data were analyzed in terms of the amount of information transmitted per second, it was found that the response time for the large matrix-matrix system had increased more than could be accounted for by the increase in the amount of information per stimulus. As has been indicated previously, the explanation of this finding seems to lie in an analysis of the coding or information-processing methods utilized by the operator. Further consideration of this problem of information processing leads to the realization that different tasks demand different orders or kinds of information processing on the part of *S*. In the systems used in these studies and in those of Fitts, *S* is simply required to relay information presented on the display onto the controls. However, in a great many operating situations the operator is required to gather information from a number of sources, integrate and interpret it, and then often simply execute a yes-no response. Although proper design of the displays and controls may facilitate his performance, it is doubtful that compatibility, in the sense of display-control interaction, would be an important variable in these latter kinds of tasks.

Examination of this question as well as an attempt to classify operational tasks on the basis of the information-processing requirements is suggested as a possibly fruitful area for future research.

Summary

The experiments reported in this paper were designed to demonstrate (*a*) interactions between three displays, light matrix, window, and auditory, and three push-button controls, matrix, row and column, and double column; (*b*) the effect of altering the size of four of these combinations upon their efficiencies; (*c*) the effect of speed upon accuracy of performance for each of the four combinations in two sizes.

The results showed:

1. There were significant interactions between the controls and displays.
2. Altering the size of the combinations did not alter the display-control relationships appreciably, although fewer stimuli per second could be handled on the larger forms.
3. However, significantly more information per second was transmitted on the smaller matrix-matrix combination. The rate of transmission of information remained constant on the two forms of the other combinations.
4. Under the forced-pacing conditions the error vs. speed trends for the various combinations were different; performance deteriorated less on the matrix-matrix combinations.
5. Presenting information at equivalent rates on the two forms of the same combination did not lead to equivalent error functions on the matrix-matrix and the window-row and column combinations.
6. Performance in the self-pacing mode of operation was superior to performance in the forced-pacing mode.

The results were discussed relative to compatibility, implications for design practice, and communication theory.

References

1. Dixon, W. J., & Mood, A. M. The statistical sign test. *J. Amer. statist. Ass.*, 1946, **41**, 557–566.
2. Friedman, M. Use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
3. Hick, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.
4. Seeger, C. M. Compatibility among the spatial aspects of stimulus and response patterns. Paper read at Midwest. Psychol. Ass., Cleveland, April, 1952.
5. Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. of Illinois Press, 1949.

(Received November 25, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

SOME CHARACTERISTICS OF THE "RANGE EFFECT"[1, 2]

CHARLES W. SLACK
Princeton University

The range effect may be described briefly as one of the sources of nonlinearity found in one-dimensional tracking of step-function inputs. For a discussion of the importance of investigating sources of nonlinearity the reader is referred to Ellson (2); for a description of the range effect see Searle and Taylor (5), Craig (1), and Slack (6). Here, it is enough to say that the range effect appears to be a tendency on the part of *S* to overshoot small inputs and undershoot large inputs in the stimulus series. Mean inputs tend to be made with no error on the average. Previous investigators have stated that this apparent regression of the responses toward the mean is not a function of the absolute amplitude of the stimuli alone but depends upon the range of preceding stimuli (4), and that individual differences are probably not significant under ordinary conditions of a small population of normal *S*s of about the same age (6).[3]

The purpose of this study is to investigate further the nature of the range effect and discover the way in which it varies as a function of speed stress. Incidentally, using these data, we are also able to test part of one of the models proposed to explain the range effect (1) and to suggest some revisions to include the present results. This experiment concerns the relationship between speed stress (frequency of stimulus presentation) and the range effect, as demonstrated by the average response-error curve and the distribution of response errors about each of several input amplitudes for each of several input speeds.

METHOD

Apparatus.—In a previous paper the author described the kind of apparatus used to produce and show range-effect response patterns (6). In brief, a Brush recording oscillograph paper feed was modified by covering up the bed with a sheet of aluminum with a slit in it. Only a short segment of the input line was visible to *S* at any time. The *S* attempted to keep a felt-pointed pen on the input line at all times and his responses were continuously recorded, superimposed upon the stimulus as in Fig. 1.

Experimental design.—In order to vary the amplitude of the stimuli in a manner which would allow the averaging of response error to inputs of a particular amplitude, an experimental design, previously described by Slack, was used (6). A "random" sequence of ten amplitudes was presented 12 times to each of a group of ten *S*s. The direction and position of the stimulus was varied from repetition to repetition within the limits set by width of the display. In this manner, the effect of past experience of *S* at each presentation of the *i*th step (where $i = 1$ to 10) was constant, provided the contribution of previous experience is expressible as a weighted mean of previous amplitudes. The distribution

[1] This study is a part of a larger project subsidized by the Office of Naval Research under their policy of encouraging basic research. The opinions expressed are those of the individual authors and do not represent the opinions or policy of the naval service. The author is also grateful to the Higgins Foundation, Princeton, for funds aiding in the purchase of research equipment.

[2] The author is indebted to Mr. Michael M. Carey for his assistance in collecting the data and preparing the report.

[3] The author has some recent unpublished data which tend to disagree with the conclusion that the range effect is indeed a function of the range of preceding stimuli. The effect of an outsized step on succeeding responses to steps of a constant amplitude falls off very rapidly indicating that, rather than being a function of the range, the range effect, a sequential phenomenon, is a function of the amplitudes immediately preceding.

of amplitudes was rectangular and ranged from .25 to 2.50 in.

The order of presentation of the amplitudes (in inches) was as follows: .50, 1.25, 1.75, 2.25, .25, 1.50, .75, 2.50, 2.00, 1.00, .50, 1.25, 1.75, 2.25, .50, 1.25, etc.

In order to vary the speed stress in a way which would control serial effects, each *S* received a different order of presentation of the frequencies according to an incomplete latin square. The frequencies used were: 1.33 step-function displacements per second, 1.00 steps per second, .66 steps per second, .50 steps per second, .40 steps per second.

Each *S*, then, was given 600 stimuli in all, 12 stimuli at each of ten input amplitudes at each of five frequencies. Of these 12, the 10 occurring latest in time were used in determining the individual mean response to the input amplitude. Any responses which were absent or questionable were discarded and it was possible to substitute one of the otherwise unused responses. Grouped data show the mean of the means of ten *S*s for ten responses to ten input amplitudes at five frequencies.

Measures.—The only response measure used was the error in the initial response amplitude as a deviation from the stimulus (see Fig. 1). The stimulus measures or control variables were: the input amplitude and the input frequency in steps per second.

Subjects and instructions.—The *S*s were ten male undergraduates from Princeton University. They were given instructions approximating those used by Ellson and Wheeler (4): "Your job is to keep this stylus on this black dot as much of the time as you can, the dot will move back and forth in this slot and you can follow it by moving the pen. Be as fast and accurate as possible." In addition, *S*s were given instructions to work in a prescribed standardized position and to keep their hands free of the machine. During a controlled practice session all *S*s were warned when they deviated from the instructions.

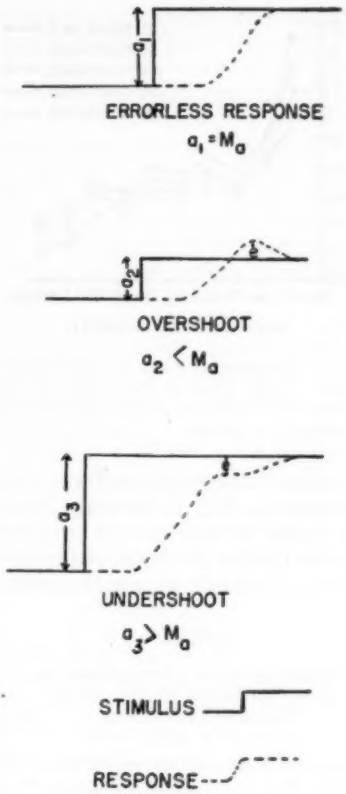


FIG. 1. Examples of responses to large, medium, and small step-function inputs. M_a refers to some weighted mean of previous inputs; e is the response error measured in this study.

RESULTS

Figure 2 shows, in grouped data, the error of the initial response to the ten step-function amplitudes at each of the five frequencies. This graph reveals that the error of the initial response appears as a family of S-shaped curves which flatten out as the frequency of stimulus input is decreased. Notice that all of the curves cross over. These group data appear to have a negative constant error which may be due to a constant tendency upon the part of *E* in scoring the tracks. The constant error appears to decrease, however, as the speed stress increases: *S*s tend to make larger responses at high frequencies of presentation. This might account for the large spread of points at .25 in. as compared with 2.50 in. The shape of the range-effect error curve leads us to believe that regression of the responses toward the mean of the dis-

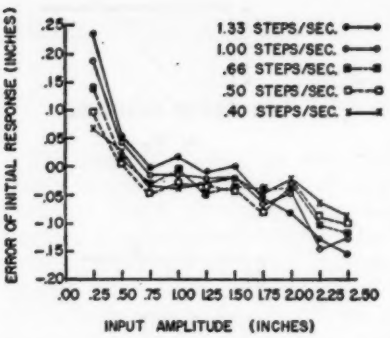


FIG. 2. Response error as a function of input amplitude. Each point is the mean of 100 responses. The parameter is speed stress or rate of presentation of stimuli.

tribution occurs in most instances only to extreme inputs. If the range effect is, in truth, a regression of extreme responses toward the mean of the distribution of stimuli, we would predict that other different sequential orders of the same step amplitudes would produce the same-shaped curve. If, on the other hand, the range effect is not independent of the sequential arrangement of the stimuli, different sequences with identical stationary characteristics should produce functions different from that in Fig. 2.[4] Table 1 shows the stationary characteristics of the distribution of response error (overshoot and undershoot) around three different step sizes. Although the mode of these distributions is difficult to establish statistically, it appears in almost all cases to be at zero error. This means that, more than they do anything else, *Ss* respond without appreciable error. It can be seen that, especially under the influence of speed stress, the distributions about extreme inputs are skewed toward the mean of the input distribution. Also it can be observed that the *SD*'s of the distributions about extreme inputs tend to increase as speed stress increases. Responses to steps of 1.50 in. (approximate mean of series) are apparently unchanged under conditions of speed stress.

TABLE 1

CHARACTERISTICS OF DISTRIBUTIONS OF RESPONSE ERROR (IN INCHES) IN TRACKING STEP INPUTS (N = 100)

Input Amplitude	Measure*	Frequency of Presentation (Steps/Sec.)		
		1.33 steps/sec.	.66 steps/sec.	.40 steps/sec.
.25 in.	Range	−.10–1.50	−.10–1.00	−.10–1.00
	Mode	.05	.00	.00
	Median	.13	.07	.03
	Mean	.23	.14	.07
	s	.33	.26	.15
	g_1	3.1	3.3	3.0
1.50 in.	Range	−.20–.35	−.45–.25	−.50–.35
	Mode	.00	.00	.00
	Median	−.01	−.03	−.02
	Mean	.00	−.03	−.02
	s	.12	.12	.12
	g_1	.58	.50	.13
2.50 in.	Range	−1.05–.30	−1.05–.30	−.75–.30
	Mode	.00	.00	.10
	Median	−.05	−.11	−.17
	Mean	−.15	−.14	.10
	s	.33	.22	.16
	g_1	−2.5	−1.5	−1.2

* The mode is very approximate since N = 100 and i = .05. The discrepancies between values of means given here and in Fig. 2 are due to grouping. The value g_1 is a measure of skewness, $g_1 = \left(\frac{\Sigma x^3}{N}\right) \cdot s^{-3}$.

DISCUSSION

Craig (1) has proposed a model to account for the range effect and for the time-delay characteristics of regressed responses which assumes that *each* response is the sum of two responses, one to the weighted mean of the distribution on the basis of previous experience and one to the presented step. The component response to the presented step occurs in Craig's model slightly later in time than the response

[4] It is suggested that, in further research on the range effect, at least six or seven different step sizes be used, since in experiments dealing with a distribution of only two or three step sizes the true nonlinear shape of the range-effect curve is obscured.

to the mean of the previous distribution. The data in Table 1 indicate conclusively that *each* response is not regressed toward the mean of the distribution by a constant amount plus or minus an error. If this were the case, all of the distributions in Table 1 would be symmetrical and under the condition of increased speed stress all measures of the central tendency would shift from zero error toward the mean of the input distribution. Since this does not happen (the distributions increasing in variance and skewness instead) we are therefore forced to reject Craig's rather ingenious hypothesis or at least to modify it in the following manner: the regressed response, i.e., the response which is the sum of two component responses, cannot be assumed to occur at all times. Instead it must be assumed that the regressed response occurs part of the time, and more often and perhaps with more regression under conditions of speed stress.

The use of three step functions in a series by Ellson and Wheeler (**4**) does not show the range effect as a curvilinear function. If we examine the group mean values for the extreme and mean step functions of the data from this experiment, the results appear similar to the results of their experiment and the range effect looks like a straight line. The sigmoid curve which results from ten step functions may indicate a possible lack of discrimination in the middle range on the part of the *S*s and/or be due to the particular sequence used. Further evidence is needed on this point. The extreme amplitudes produce distributions of response errors which are skewed toward the mean input amplitude. The input amplitude close to the mean produces a more symmetrical distribution. The only conclusive change in the distributions of error of initial response with increase in speed stress appears to be an increase in the variance, skewness, and hence the means toward the mean of the input distribution. With regard to the parameter of frequency of stimulus presentation or speed stress, the data illustrate that the range effect increases as the frequency increases. We would expect that at high frequencies the range effect would be at its greatest just before the response pattern disintegrates. On the other hand, if the frequency is decreased sufficiently, it is suggested that the range effect would approach zero as a minimum.

A proper evaluation of the range effect as a source of nonlinearity in the human operator awaits further research at extreme frequencies and with other types of stress-inducing conditions.

Summary

This investigation was undertaken in order to discover the shape of the range-effect error curve and the shape of the distributions of response errors under conditions of varying speed stress resulting from increasing stimulus frequency.

An experimental design was used which controlled the previous experience of *S* before each stimulus presentation so that an averaging of otherwise time-dependent data could be made.

1. The shape of the range-effect curve for a particular sequence of stimuli was found to be **S** shaped on the average and not linear.

2. The distributions of error at each input amplitude were found to have modes near zero and to be skewed toward the mean of the input distribution.

3. Under conditions of lower speed stress the range effect decreases, the

error curves flatten out, the *SD*'s of the distributions of response errors to extreme inputs decrease and the distributions tend to be less skewed.

A model for the range effect proposed by Craig (1) was discussed in the light of these results.

References

1. Craig, D. R. Effect of amplitude range on duration of response to step-function displacements. USAF Air Materiel Command, Wright Field, Dayton, Ohio, September, 1949. *Technical Report* No. 5913.
2. Ellson, D. G. The application of operational analysis to human motor behavior. USAF Air Materiel Command, Wright Field, Dayton, Ohio, May, 1948. Serial No. MCREXD-694-25.
3. Ellson, D. G., Hill, H., & Craig, D. R. The interaction of responses to step-function stimuli: II. Equal opposed steps of varying amplitude. USAF Air Materiel Command, Wright Field, Dayton, Ohio, August, 1949. *Technical Report* No. 5911, Serial No. MCREXD-694-2P.
4. Ellson, D. G., & Wheeler, L., Jr. The range effect. USAF Air Materiel Command, Wright Field, Dayton, Ohio, May, 1949. *Technical Report* No. 5813, Serial No. MCREXD-694-2P.
5. Searle, L. V., & Taylor, F. V. Studies of tracking behavior. I. Rate and time characteristics of simple corrective movements. *J. exp. Psychol.*, 1948, **38**, 615–631.
6. Slack, C. W. Learning in simple one-dimensional tracking. *Amer. J. Psychol.*, 1953, **66**, 33–44.
7. Taylor, F. V., & Birmingham, H. P. Studies of tracking behavior. II. The acceleration pattern of quick manual corrective responses. *J. exp. Psychol.*, 1948, **38**, 783–795.

(Received December 4, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

PSYCHOLOGICAL PROBABILITY AS A FUNCTION OF EXPERIENCED FREQUENCY[1]

FRED ATTNEAVE

Perceptual and Motor Skills Research Laboratory, Human Resources Research Center

The perception of objects under conditions of stimulus impoverishment may be analyzed in terms of concepts which Shannon (**9**) has developed for dealing with the formally similar problem of communication over a noisy channel. Perception with better-than-random accuracy is possible under such conditions because (*a*) different events (e.g., objects, or parts or aspects thereof) have different probabilities of occurrence, and (*b*) these probabilities are modified by context to the extent that in many cases they become practical certainties. Numerous tachistoscopic experiments (**10,** p. 737 ff.) have demonstrated that words may be read correctly despite considerable mutilation. More recently, Howes and Solomon (**5**) have found correlations ranging from −.68 to −.75 between the visual-duration thresholds of words and the logarithms of their relative frequencies as determined by the Thorndike-Lorge counts. Both of these lines of evidence are consistent with the view that a sufficiently high probability may serve as a substitute for actual stimulation in the perceptual process. James (**6**) summed the matter up crudely but aptly with his observation that "perception is of definite and probable things."

It is obvious, however, that environmental probabilities (i.e., the relative frequencies with which various events actually occur in various contexts) can influence perception *only to the extent that they are appreciated and utilized by the observer.* It likewise follows from the above considerations that the accuracy of perception under adverse circumstances will depend in large measure on the observer's appreciation of relevant environmental probabilities. Comparatively little has been done toward determining how closely psychological probabilities correspond to their environmental counterparts, or by what means and with what difficulty such correspondences are established and improved by experience, though some well-directed attacks on these problems have been made by Brunswik and his associates (**1, 2**), and by Preston and Baratta (**8**).

In the present study, direct comparisons were made between environmental and psychological probabilities for the letters of the alphabet. The letters were chosen for study, in preference to any other class of objects, because they present a unique combination of properties: (*a*) Their relative frequencies are extremely stable (from sample to sample of English text); (*b*) they are observed in these proportions by most of the members of our culture throughout their entire lives, so that whatever probability-learning occurs under such circumstances may be supposed to have

[1] The experimental work for this study was performed as part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

The author wishes to acknowledge the assistance of S/Sgt. Richard B. Lake, S/Sgt. Herbert E. Heldt, and A/1C Noel C. Kruger, in carrying out this study.

reached a practical limit by adulthood. The psychological probabilities of the letters were assessed by two general procedures. In the first, *Ss* were required to make direct quantitative judgments of relative frequency. In the second, the relative frequencies with which *Ss* guessed letters in a pseudo-ESP experiment were determined. The latter procedure was carried out under three different sets of instructions designed to vary bias with respect to relative probabilities. True or environmental frequencies were obtained from samples of English text by direct count.

Method

Subjects.—A total of 204 airmen basic trainees at Lackland Air Force Base served as *Ss*. Of these, 90 constituted the frequency-judgment group; the remaining 114 were divided among three guessing groups of 38 each.

Procedure for obtaining judgments.—Each *S* was given a mimeographed sheet on which the alphabet was presented in a single column with a blank by each letter. Instructions were as follows:

"This is a test to find out how observant you are.

"All of you read newspapers, and magazines, and books, from time to time. Now, in ordinary reading matter of this sort, some of the letters of the alphabet are used much more often than others. We want to find out how well you have noticed which letters of the alphabet are used most often, and which ones are very seldom used.

"Suppose you took a newspaper clipping with a total of a thousand letters in it. This clipping I'm holding up has just about a thousand letters in it. Suppose you counted all the A's in this clipping, then all the B's, then all the C's, and so on down through the alphabet. How often do you think each letter would turn up? This is what you are to fill in on the blanks I have given you: how often, on the average, you think each letter would appear in a sample of reading matter about this long, with a total of a thousand letters in it.

"The estimates you put down actually ought to add up to a thousand, but don't worry about trying to make them add up—the important thing is to get the proportions of letters to each other as nearly right as you can. I will point out this, though: that if all the 26 letters were used equally often, each one would appear about 38 times. Therefore, if you think a particular letter is used *more* often than average your guess should be *above* 38, whereas if you think it is used *less* often than average your guess should be *below* 38.

"You will have only 10 min. in which to make these estimates, so work as quickly as you can. Are there any questions before we start?"

The statement concerning a time limit was inserted to encourage spontaneous and unstudied judgments. Actually no time limit was adhered to, but few *Ss* required more than 10 min.

Guessing procedure.—In this procedure, *Ss* attempted to "perceive" a series of 100 letters which in reality were completely nonexistent. Three groups of *Ss* were employed. Group A was given no indication of the relative frequencies of the fictitious letters to be guessed; Group B was led to believe that they occurred with approximately uniform frequency; and Group C was led to believe that they occurred with English text frequencies. The following applies to all three groups, except as otherwise indicated.

Each *S* was given mimeographed sheets with blanks numbered from 1 to 100. Before reading the instructions, *E* placed a small sealed package (containing blank slips of paper of approximately playing-card size) in a conspicuous position on the table before him. Instructions were as follows:

"In the test you are about to take, all you will have to do is to guess. Here on the table is a sealed package with 100 small cards in it. Each one of these cards has some letter—just a single letter—printed on it. These cards, with their letters, have been thoroughly shuffled, so that they are not in any special order. You are going to try to guess, one by one, what these letters are—without ever seeing the cards."

(The *Ss* were then assured that the experiment had significance even though the task appeared to be impossible, and were urged to follow the instructions as carefully as possible.)

"First imagine yourself taking the wrapper off the cards, and looking at the top card. Try to imagine—or guess—what letter is printed on that card. Then imagine that you turn over the top card, and look at the second card. Try to imagine what letter is printed on the second card, and so on down through the pack of 100.

"You will print plainly the letter you guess for the top card on Line 1 of your answer sheet, then print on Line 2 your guess for the card second from the top, and so on down. I've given each of you a folded sheet of blank paper. Slide this down over your answer sheet as you go along, so as to cover up everything you have just written: the idea of this is to keep the letters you

have already written down from influencing your guesses.

"Now, as you do this, don't try to think too hard about what you are doing. It will be better if you can just make your mind a blank each time you guess, and put down the first letter that pops into your head.

"If you still don't understand what you are supposed to do, I'll try to answer any question you have, but don't ask your question out loud before the group. Just raise your hand if you have a question, and when I come to your seat talk in as low a voice as you can.

"Go ahead whenever you are ready."

Group A received only the instructions above. An additional paragraph was included in the instructions to Groups B and C, just before the paragraph beginning "If you still don't understand. . . ." For Group B this was as follows:

"You may be interested in knowing how the pack of cards here was made up in the first place. *The letters were drawn at random out of a hat, and after each draw the letter was returned.* After a hundred cards were printed with letters drawn in this way, they were thoroughly shuffled before they were sealed up. This means that the letters here may be in any order, that any letter may come up any number of times, and that *for each guess you make, any letter is just as likely to be right by chance as any other letter.*"

Group C had the same instructions as Group B except that the italicized portions were replaced by the following: "*The letters were chosen by going through a newspaper article and copying every fifth letter off on to one of the cards. . . . on the average any particular letter will come up just about as often in the pack here as it does in ordinary reading matter.*"

The guesses of all *S*s were carefully inspected for meaningful, familiar, or otherwise ordered sequences of letters. Where any evidence of such serial dependencies was found (e.g., sequences which spelled out words, sequences in alphabetical order, regular alternations, etc.), the guesses of the *S* in question were eliminated from further consideration. Additional eliminations were made randomly to equate the *N*'s of the three groups. Thus each group ended with 38 out of an original 45 *S*s.

Letter-frequency count.—The frequency with which each letter of the alphabet occurred was tabulated for samples of English text containing a total of 10,341 letters. Of these, 5170 came from articles in a San Antonio daily newspaper, and 5171 from articles and stories in popular magazines. The coefficient of correlation between newspaper and magazine frequencies (.978) was subsequently used as a split-half reliability coefficient.

TABLE 1

TRUE, JUDGED, AND GUESSED FREQUENCIES OF LETTERS

Letter	Frequency				
	English Text	Judged (Mdn)	Guesses		
			Group A	Group B	Group C
A	845	75	207	216	248
B	164	45	138	173	149
C	291	45	213	168	191
D	424	40	157	147	178
E	1270	60	152	134	244
F	223	35	142	124	134
G	256	32	112	108	109
H	606	35	119	100	140
I	729	45	135	141	141
J	14	25	122	92	92
K	92	20	127	108	114
L	392	35	177	170	152
M	258	38	158	169	178
N	780	38	154	163	176
O	774	45	214	211	198
P	199	30	166	175	146
Q	10	10	119	127	54
R	645	40	153	170	206
S	623	40	123	163	166
T	925	44	182	188	244
U	281	30	82	79	75
V	96	15	105	125	103
W	228	30	128	127	133
X	18	10	137	133	70
Y	185	25	125	137	99
Z	13	9	153	152	60
Sum	10,341	896	3800	3800	3800
SD	329.9	14.7	32.2	33.8	54.7
$\frac{1000 \times SD}{\text{Sum}}$	31.9	16.4	8.5	8.9	14.4

RESULTS

The data obtained are summarized in Table 1. The English-test frequencies listed are totals for the sample of 10,341. The judged frequencies are median ratings, from 90 *S*s, of occurrences-per-1000 letters. The guessed frequencies are totals for groups of 38 *S*s, each of whom made 100 guesses. The statistic $\frac{1000 \times SD}{\text{sum}}$ expresses SD in terms of occurrences-per-1000 letters, and thus constitutes a measure of dispersion which is comparable among all the columns.

A split-half reliability coefficient was computed for each response vari-

TABLE 2

INTERCORRELATIONS AMONG THE VARIABLES*

Variable	English-Text Frequency	Judged Frequency (Mdn)	Guessing Frequency Group A	Guessing Frequency Group B	Guessing Frequency Group C	Alphabetical Order
English-text frequency	.98 (.99)	.79 (.81)	.42 (.45)	.43 (.46)	.83 (.86)	−.30
Judged frequency (median)	.79 (.81)	.96 (.97)	.55	.54	.88 (.93)	−.65
Guessing frequency, Group A	.42 (.45)	.55	.76 (.86)	.85	.69	−.33
Guessing frequency, Group B	.43 (.46)	.54	.85	.78 (.88)	.67	−.13
Guessing frequency, Group C	.83 (.86)	.88 (.93)	.69	.67	.87 (.93)	−.46
Alphabetical order	−.30	−.65	−.33	−.13	−.46	1.00

* Parenthetical figures are Spearman-Brown corrections in the case of reliability coefficients, and corrections for attenuation in the case of correlation coefficients.

able by correlating measures based on a random half of the *S*s with corresponding measures based on the other half. These coefficients, both uncorrected and corrected by the Spearman-Brown formula, appear on the diagonal of Table 2. A discussion of this relatively unfamiliar use of the Spearman-Brown formula may be found in Guilford (**3**, p. 421).

Table 2 further presents intercorrelations among all the variables studied. Significance levels for these coefficients (based on an N of 26) are as follows: for $r = .39$, $p = .05$; for $r = .50$, $p = .01$; and for $r = .61$, $p = .001$. Thus all the coefficients of correlation in the table, except some of those in which alphabetical order is a variable, are significant beyond the .05 level of confidence. It should be pointed out that the statistical procedures here employed apply to letters the treatment more usually given to *S*s, and hence test what Brunswik (**1**) has called *ecological* significance. It is considered that the issue of *population-al* significance may be ignored with reasonable safety, in view of the high inter-*S* reliabilities obtained.

Several comments upon the results may be made:

1. Guesses of Group C (in which *S*s were told that the letters were derived from English text) are as highly correlated with English-text frequencies as are direct judgments thereof. Each of these relationships was found to have a moderate degree of curvilinearity which was corrected by a log-log plot. When correlation coefficients were comp ted on this basis, a coefficient of .88 (uncorrecte) was found between log true frequency and log judged frequency, and a coefficient of .86 between log true frequency and log frequency of guesses in Group C. It appears, therefore, that in each of these two cases about three-quarters of the response variance is attributable to true (English-text) frequency.

The goodness of the relationship between judged and true frequency does not imply, however, that judgments were highly *accurate;* it means rather that errors were of a lawful nature. Low frequencies were systematically overestimated and high frequencies were systematically underestimated: thus it may be seen in Table 1 that the *SD* of judged frequencies is only about half that of true frequencies, when the two measures are expressed in comparable terms. The guessing frequencies of Group C closely resemble judged frequencies in dispersion.

2. The possibility that alphabetical order might have served to some extent as a mediating variable in the relationships between objective and psychological frequencies was considered, but does not find support in the data. Holding the effect of order constant by partial correlation results in only trivial alterations of coefficients.

3. The differences in set among the three guessing groups resulted in a marked contrast between the guessing behavior of Groups A and B on one hand and that of Group C on the other. Groups A and B agree with each other about as closely as their reliabilities permit (see Table 2).

Only about a fifth of their variance is attributable to English-text frequency, whereas nearly three-quarters of the variance of Group C may be so explained; this difference proves highly significant when evaluated by the formula for correlated coefficients of correlation (7, p. 124–125, Equations 47 and 49). Thus, instructing *Ss* that the fictitious letters would occur with English-text frequency appears to have influenced their guesses markedly in the expected direction; whereas instructing them that the letters would occur with uniform frequency seems to have had no effect whatever.

It should be pointed out that the frequencies yielded by each of the three groups show a closer relationship to judged than to true letter frequencies. It is further apparent, from the reliabilities of the groups, that guessing frequencies show a degree of inter-*S* consistency far in excess of that attributable to English-text frequency. The author has been so far unable to determine the source of this consistency.

Discussion

One of the few other studies that have undertaken to investigate the psychophysics of probability is that of Preston and Baratta (8). Since their techniques were different from ours in almost every respect, certain aspects of agreement in results suggest that principles of considerable generality are involved. Preston and Baratta found that *Ss* in a gambling situation tended to make unrealistically high bets when the objective probabilities were low, and unrealistically low bets when the true odds were high. The "indifference point," at which psychological and mathematical probabilities were equal, occurred near the geometrical mean of the series ($p = .24$). The authors suggest an explanation in terms of the adaptation-level theory of Helson (4).

It may be seen in Fig. 1 that *Ss* in the present study likewise overestimated low probabilities and underestimated high ones. Here, however, the "indifference point" is very close to the *arithmetical* mean. The fact that the value of the arithmetical mean was explicitly mentioned in the instructions may have some importance in this connection.

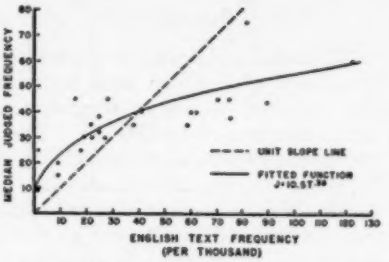


FIG. 1. Median judged frequency plotted against English text frequency. Each point represents a letter of the alphabet.

A linear-regression function was fitted by the least-squares method to the log-log plot corresponding to Fig. 1. Translated out of logarithmic terms, it becomes a power function with the equation $J = 10.5T^{.36}$. If one were willing to make the Fechnerian assumption that psychological magnitudes are proportional to summated j.n.d.'s, this equation might be differentiated and stated in the form of Guilford's "generalized n^{th} power law" (3, p. 138), thus: $\Delta T = kT^{1-.36}$. The value of n so obtained ($n = 1 - .36 = .64$) has interest because it is fairly typical of values which have been obtained in conventional psychophysical experiments.

An interpretation of the equation $J = 10.5T^{.36}$ in terms of adaptation-level theory might assume that judged frequency is a weighted geometrical average of true frequency and adaptation level, i.e., that $J = A^n T^{(1-n)}$. It would follow that $A^{.64} = 10.5$, and hence that $A = 39$. This value, which is slightly greater than the arithmetical mean, represents the "indifference point" at which letter frequencies are neither overestimated nor underestimated.[2]

[2] A complementary point of considerable interest is that Guilford's "n^{th} power law" may be

Summary

Numerical judgments of the relative frequency with which each letter of the alphabet occurs in English text were obtained from 90 *Ss*. Three additional groups, each containing 38 *Ss*, attempted to guess the identities of 100 fictitious letters in a pseudo-ESP experiment. Group A was given no bias as to the relative frequencies of the letters, Group B was told that they occurred with uniform frequency, and Group C was told that they occurred with English-text frequency, though in a disordered sequence.

The correlation between median judged frequency and "true" English-text frequency (determined by direct count), was found to be .79 without transformation, and .88 after a log-log transformation. The frequency with which each letter was guessed by Group C was correlated with English-text frequency to an equivalent degree (untransformed, $r = .83$; log-log, $r = .86$). Groups A and B yielded correlations with true frequency in the low .40's (still significantly greater than zero).

Intersubject consistency within the guessing groups was much higher than agreement with English-text frequency could explain.

Results are compared with those of Preston and Baratta (**8**), and discussed in terms of the psychophysical formulations of Guilford (**3**) and Helson (**4**).

derived from adaptation-level theory. One of Helson's basic assumptions is that the j.n.d. is proportional to some average of stimulus value and adaptation level, i.e., that

$$\frac{\Delta R}{f(R, A)} = K.$$

Although he has used an arithmetical average in certain of his studies, Helson considers that a weighted geometrical average may be somewhat more appropriate (personal communication). In this case the equation becomes

$$\frac{\Delta R}{A^{1-n}R^n} = K.$$

But since A^{1-n} is a constant, it may be combined with K, giving us

$$\frac{\Delta R}{R^n} = K'$$

which is Guilford's "generalized n^{th} power law" (**3**).

References

1. Brunswik, E. Systematic and representative design of psychological experiments. In J. Neyman (Ed.), *Proceedings of the Berkeley Symposium on mathematical statistics and probability*. Berkeley and Los Angeles, Calif.: Univer. of Calif. Press, 1949. Pp. 143–202.
2. Brunswik, E., & Herma, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, **41**, 281–290.
3. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
4. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
5. Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41**, 401–410.
6. James, W. *The principles of psychology*. New York: Holt, 1890.
7. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
8. Preston, M. G., & Baratta, P. An experimental study of the auction-value of an uncertain outcome. *Amer. J. Psychol.*, 1948, **61**, 183–193.
9. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, 379–423, 623–656.
10. Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received November 25, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

EFFECT OF AMOUNT OF VERBAL ANCHORING AND NUMBER OF RATING-SCALE CATEGORIES UPON TRANSMITTED INFORMATION

A. W. BENDIG AND J. B. HUGHES II
University of Pittsburgh

The conceptual syntax of Shannon's information theory (6) has stimulated new mathematical analyses of old psychological problems. Recently Garner and Hake (2) have extended the application of information theory to judgmental rating scales. Their analysis suggests that an informational approach to the problems in constructing and refining rating scales may offer advantages over the more traditional psychometric methods of assessing the reliability and validity of such scales. The present study was concerned with the application of information theory to the construction of scales measuring self-ratings.

Probably the first problem that faces the constructor of a rating scale is the number of categories to include on the scale. The scale should not be too coarse or some of the discriminative ability of the rater is lost. Nor should the number of categories greatly exceed the ability of the rating to discriminate differences between the stimuli and so introduce added error variance to the ratings. A second problem is how much the scale categories should be verbally anchored by descriptive phrases or examples. Traditionally, the more verbal anchoring used, the more adequate are the ratings, but in specific cases, such as in self-ratings where the occurrence of "projective" phenomena may increase the validity of the ratings, such a principle may not be applicable. Within the context of information theory these questions can be phrased as the effects upon transmitted information of variations in methods of verbally anchoring the response categories.

Two studies have reported some findings on the functional relationship of number of response categories and transmitted information. Hake and Garner (4) found that the absolute amount of information transmitted by observers' ratings of the pointer on an instrument dial rose appreciably when the number of response categories was increased from five to ten, but increased only slightly for any increases above ten. Pollack (5) reports that the transmitted information of encoded messages increased concomitantly with increases in the number of response alternatives. However, little information as to experimental procedure and the range of categories used was given in the only available report. These two studies suggest that transmitted information increases as a function of increases in the response alternatives available to the rater when the rated stimuli are objectively observed visual and auditory stimuli. Whether this conclusion is descriptive of the results found when the raters are assessing their own subjective feelings forms the major question stimulating the present study.

The study reported below was designed to measure the effect upon the transmitted information in self-ratings of variations in two rating-scale parameters: (*a*) the number of categories on the scale, and (*b*) the kind of anchoring used to verbally define the scale categories.

PROCEDURE

Scales.—Fifteen different numerical rating scales were constructed using five different numbers of rating-scale categories (3, 5, 7, 9, and 11 categories) and three conditions of verbal anchor-

ing of the categories (center category anchored, the two end categories anchored, and both center and end categories anchored). The verbal anchoring statements used were:

I know a great deal about this country.
I know something about this country.
I know very little about this country.

The stimuli to be rated were the names of 12 foreign nations, ranging from familiar nations such as Canada and France to lesser-known countries like Sweden and Egypt. The lowest category on each scale was given a numerical value of 1 and the top category a value of 3, 5, 7, 9, or 11, depending on the particular scale. Thus the scales were "single" scales in the sense used by Guilford (3, p. 268). The particular scale, stimuli, and instructions were mimeographed on the single sheet given to *Ss*.

Subjects.—The *Ss* were 225 college students enrolled in introductory and social psychology classes. Each of the 15 scales was used to rate the stimuli by 15 randomly selected *Ss*. The *Ss* were instructed to rate the stimuli as to how much *S* knew about the geographical, political, economic, and sociological characteristics of each of the 12 foreign countries.

Results

The information analysis of the data followed the computational procedure described by Garner and Hake (2). Matrices were constructed with the 12 stimuli (names of foreign countries) along one axis, scale responses along the other axis, and the proportion of *Ss* rating each stimulus at each scale category constituting the cell entries (2, p. 449). Separate analyses were computed on each of the 15 subgroups. In addition, the three subgroups using different numbers of verbal anchors were combined to achieve an over-all analysis for each number of scale categories. The results of these analyses are in Table 1. In each analysis reported in the table, stimulus information had a magnitude of 3.58 bits.

In comparing the effect upon information of variations in the number of scale categories, mean measures of response information, response equivocation, and transmitted information were computed for each of the category groups and are shown in Table 2. Also given in this table are the same information measures computed when the three anchor groups for each category were combined. The differences between these pairs of values represent gains or losses in information due to the combining of groups with different rating orientations resulting from differing conditions of scale anchoring. It can be seen that mean transmitted information increases rectilinearly as a function of increasing numbers of scale categories. However, when anchor groups are combined, this increase is not very evident. Table 2 indicates that combining anchor groups increases response equivocation, which lessens transmitted information in the combined groups.

In Table 3 are given mean information measures for each of the anchoring conditions. As expected, transmitted information increases with increased verbal structuring of the scale, but the magnitude of increase is not very marked.

Discussion

The results of this study support those reported by Hake and Garner (4). Transmitted information increases as more categories are added to the rating scale when either objective, external or subjective, internal stimuli are being judged. Hake and Garner used 5, 10, 20, and 50 categories of response and found little increase in transmitted information above 10 scale categories. Our study investigated the range of categories represented only by their two lowest steps and found an essentially rectilinear increase. However, there appears to be a slight deceleration from Steps 9 to 11 in the increase in transmitted information, which may reflect the diminishing returns noted by Hake and Garner. One point to be noted is the much smaller magnitude of transmitted information found in those self-ratings, averaging about one-sixth that found for objective stimuli.

The differences noted in Table 2 between mean information measures

TABLE 1

INFORMATION MEASURES FOR RATING SCALES

Number of Categories	Response Information				Response Equivocation				Transmitted Information			
	Number of Anchors				Number of Anchors				Number of Anchors			
	1	2	3	All	1	2	3	All	1	2	3	All
3	1.49	1.52	1.36	1.49	1.13	1.20	.95	1.22	.36	.32	.41	.27
5	2.19	2.18	2.27	2.22	1.87	1.77	1.76	1.97	.32	.41	.51	.25
7	2.58	2.76	2.66	2.70	2.16	2.28	2.10	2.41	.42	.48	.56	.29
9	3.06	3.02	3.05	3.06	2.52	2.48	2.49	2.74	.54	.54	.56	.32
11	3.37	3.37	3.40	3.40	2.81	2.75	2.81	3.08	.56	.62	.59	.32

TABLE 2

COMPARISON OF MEAN INFORMATION MEASURES FOR SUBGROUPS AND SIMILAR MEASURES WHEN ANCHORING SUBGROUPS ARE COMBINED

Comparison	Scale Categories				
	3	5	7	9	11
Response information					
Combined	1.49	2.22	2.70	3.06	3.40
Mean (subgroups)	1.44	2.21	2.67	3.04	3.38
Difference	0.05	0.01	0.03	0.02	0.02
Response equivocation					
Combined	1.22	1.97	2.41	2.74	3.08
Mean (subgroups)	1.09	1.80	2.18	2.50	2.79
Difference	0.13	0.17	0.23	0.24	0.29
Transmitted information					
Combined	0.27	0.25	0.29	0.32	0.32
Mean (subgroups)	0.36	0.41	0.49	0.55	0.59
Difference	−0.09	−0.16	−0.20	−0.23	−0.27

TABLE 3

MEAN INFORMATION MEASURES FOR EACH ANCHORING CONDITION

Measure	Scale Anchoring		
	Center	Ends	Center and Ends
Response information	2.54	2.57	2.55
Response equivocation	2.10	2.10	2.02
Transmitted information	0.44	0.47	0.53

averaged over the subgroups differing in scale anchoring and the same information measures computed when these subgroups were combined appears attributable to the subjective judgmental scaling used by the raters within each subgroup. Volkmann (7) has recently reviewed the experimental evidence of the effect upon raters' judgments of variations in anchoring conditions used with rating scales. Variations in the placement of anchoring stimuli upon the scale used to rate other stimuli result in shifts in both the central tendency and variability of rater judgments. It has been shown in a previous paper (1) that increased anchoring of our scales somewhat increased interrater reliability in judging the stimuli. Differences between scales in the number of anchors used will result in (*a*) differences in the mean ratings assigned the stimuli by groups using different scales, and in (*b*) greater or lesser agreement among the raters using any one of the scales. Combining subgroups whose mean stimuli ratings

differ markedly would be expected to increase response equivocation, as was shown in Table 2. This explanation implies that the response equivocation computed on the combined groups is composed of two parts: (*a*) the sum of the response equivocation *within* each of the three anchoring subgroups that have been combined, and (*b*) the constant differences *between* subgroups in mean ratings assigned the 12 stimuli. An analogy is that of the simple analysis-of-variance procedure where the total variation in a series of measurements is split into the orthogonal components of random variation within groups and constant variation between groups.

One practical outgrowth of our results is a comparison of the suggestions that could be made to the constructor of a rating scale from the results of the previous and more traditional analysis of the reliability of these scales (1) and the present analysis within an informational syntax. In the reliability analysis no increase in reliability was found when the number of scale categories was increased from 3 to 9 and a slight decrease in reliability with 11 scale categories. Therefore we concluded that 3 categories could be used just as well as 9, and that 11 or more categories should not be used. The present results indicate that 9 categories will yield more transmitted information than any smaller number of scale categories and should be recommended. In fact, more transmitted information was obtained with 11 categories, although 11 categories were less reliable. It is interesting to speculate whether the deceleration in the increment of transmitted information gained by increasing scale categories from 9 to 11 is related to the decrease in scale reliability over this interval.

As to the number of anchors, both analyses (psychometric and informational) suggested the superiority of increased anchoring. Both the reliability and information transmitted by the scales increased as functions of additional anchoring.

Summary

Subjects ($N = 225$) rated themselves on their knowledge of 12 foreign countries on rating scales with 3, 5, 7, 9, or 11 categories and with these scales verbally anchored either in the center, at both ends, or at both center and ends. The data were analyzed within an information theory syntax as to the effect of variations in number of scale categories and amount of verbal anchoring upon the information transmitted by the scale. Results indicated an increase in the absolute amount of transmitted information as the number of scale categories was increased. Increased verbal anchoring of the rating scale resulted in a slight increase in the information transmitted by the scale.

References

1. Bendig, A. W. The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories on the scale. *J. appl. Psychol.*, 1953, **37**, 38–41.
2. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
3. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
4. Hake, H. W., & Garner, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. *J. exp. Psychol.*, 1951, **42**, 358–366.
5. Pollack, I. The assimilation of sequentially encoded information. *Amer. Psychologist*, 1952, **6**, 266–267.
6. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1945, **27**, 379–423, 623–656.
7. Volkmann, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer and M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951. Pp. 273–294.

(Received December 8, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

TWO-CHANNEL LISTENING[1]

E. C. POULTON

Medical Research Council Applied Psychology Research Unit, Cambridge, England

The task of selecting occasional relevant pieces of information during prolonged listening to two simultaneous verbal sources, so far appears to have escaped experimental study. The present aim was to investigate in this situation the effects and interactions of four independent variables: distribution of attention, separation of speech sources, density of communication, and similarity of relevant information to irrelevant.

METHOD

Speaker separation experiment.—Information was presented simultaneously over two loudspeakers, using a double-track Ferrograph magnetic tape recorder. One-half of each experimental period used speakers opposite right and left ears, with approximately 180° horizontal angular separation; the other speakers in front with a 3½–6° vertical separation (one placed on top of the other). The tracks represented aircraft–control-tower communication channels, the mean intensity of each being about 75 db when it reached *S*.

The *S* had to listen either to one speaker (restricted attention) or to both (distributed attention), for aircraft calling one particular tower. Calls to three other towers had to be neglected. When he heard such a call, he had to write down the aircraft number, which always had three digits, the speaker number over which the call came, and the time in minutes and seconds to the nearest 5 sec. He was asked to guess the number of any aircraft only partly heard. If he failed to hear an aircraft number at all, but thought it was calling his tower, he had to leave a blank, simply writing down the speaker number and time.

During each experimental period of 1 hr., two kinds of information were presented, all by the same two voices: (*a*) 64 calls to each tower from visiting aircraft, scattered throughout the hour, e.g., "Lakenheath tower, 123, over"; and (*b*) 15-min. continuous blocks of standard Royal Air Force pattern conversations between each tower in turn and three aircraft in its vicinity, e.g., "Lakenheath tower, 210, taxi clearance, over"; "210, Lakenheath tower, clear to taxi, runway 25, Queenie Nan How 1003 millebars, 30 feet, over"; "Lakenheath tower, 210." During blocks of conversations involving *S*'s tower, which occurred once in each half of each experimental period, he had only to write down the numbers of each of the three aircraft on the first occasion upon which he heard them. He did not have to record all the subsequent conversations made by them.

The experiment was performed by 16 *S*s, four at a time. They sat at tables, separated by screens. They received a 1¼-hr. instructional and practice period on one day, using material similar to that used in the experiment, and four experimental periods each of about 1-hr. continuous listening on the following days. The material presented was always the same, but each *S* was allocated a different tower each day. At the start of each period *S*s were told their previous day's results. Practice and transfer effects both within periods and from one period to the next were equated by græco-latin-square and balanced-group designs.

Communication density experiment.—In this the density of communication was varied, while the position of the speakers was held constant. They were placed nearly symmetrically front right and front left of *S*, with a horizontal angular separation of 50–55°. In 15 min. of each experimental period the silence was broken at intervals by single calls from visiting aircraft over either speaker to any one of the four towers. In two further 15 min. a continuous block of conversations to one of the towers, interspersed with calls to any of the towers from visiting aircraft, was presented over one speaker, while calls from visitors were presented singly over the other speaker. And in the fourth 15 min. a continuous block of conversations, interspersed with calls, was presented over both speakers simultaneously. Conditions were otherwise similar to those of the previous experiment, 16 new *S*s being used.

Directed attention experiment.—Sixteen additional *S*s were given the same 1¼-hr. instructional and practice period as in the speaker sep-

[1] This research has been under the general direction of Professor Sir Frederic Bartlett F.R.S. and of Dr. N. H. Mackworth. The author was receiving financial support from the British Medical Research Council. The *S*s were supplied by the Royal Navy. The close cooperation of Mr. D. E. Broadbent is gratefully acknowledged.

aration experiment on one day, and then received a single 1-hr. experimental period the next day. In this they had to listen for 36 calls from visiting aircraft, which the results of the two previous experiments had shown to have given difficulty. These calls were all selected from the high density quarter of the communication density experiment, the material of which was identical with that of one-quarter of the speaker separation experiment.

Before each call *S* wrote down the channel to which to listen, the tower to which the call was to be made, and the numbers of any other aircraft calling this tower on this channel before the call in question. He was also told the order of communications to expect on this channel before the call. The two-channel record was started one conversation (15–30 sec.) before the call, except in eight cases in which two selected calls on different channels were separated on the tape by a shorter time interval, when only this time was allowed for the second. In the four cases in which two calls followed each other closely on the same channel, both had to be listened for in the same trial. Half the *S*s received the first 18 calls with the speakers opposite right and left ears, and the second 18 with the speakers together in front. The other half received the calls with the reverse arrangement of speakers.

Subjects.—The 48 *S*s were all enlisted men in the British Royal Navy, aged between 17 and 29. The groups performing the three experiments were of approximately similar composition.

Scoring.—The loud-speaker number and time of each call were used simply to determine the signal to which *S* was responding. Three types of aircraft number errors were differentiated: (*a*) *omission;* (*b*) *mishearings,* when an incorrect number was recorded at the correct time, and generally attributed to the correct loudspeaker (only when the speakers were together was it difficult to tell from which a particular call came); (*c*) *false selections,* when no call was made to that tower at that time.

Calculations.—Analysis of variance was not attempted, owing to the large number of zero component scores on breaking down the over-all scores. The significance of the difference of each pair of means was assessed separately (*a*) against the variability of the *S*s for all the calls represented in the means combined, and (*b*) against the variability of the individual calls for all the *S*s combined. Differences have been accepted as significant only when both criteria were satisfied. Where the two levels were different, the less significant was always taken. This method of assessing significance is not as sensitive as analysis of variance. Significance levels beyond $p = .001$ will be called "highly significant," beyond $p = .01$ "significant," and beyond $p = .05$ "just significant." Both tails of the assumed normal distribution have been used except where stated.

Results

1. In all three experiments omissions and mishearings were distributed independently of each other. Further, only 2.8% of omissions were correctly noted, although *S* was instructed to report all of which he was aware; and on being told his previous day's results, he often reported spontaneously that he did not realize he had missed any calls.

2. Three types of masking can be distinguished in the communication density experiment: (*a*) calls from visiting aircraft *surrounded* on both sides by conversation on the same channel; (*b*) separated calls on one channel *covered* by continuous conversation on the other; (*c*) *surrounded* calls on one channel *covered* by conversation on the other. Surrounded calls gave significantly more omissions than mishearings, while covered calls gave just significantly (one tail) more mishearings. The full results are shown in Table 1.

TABLE 1

Masking Effects

Type of Masking	Percentage Errors	
	Omissions*	Mishearings**
Unmasked	.4%	.2%
Surrounded	4.0%	1.5%
Covered	1.4%	3.9%
Surrounded and covered	8.7%	4.7%

* All differences significant at .01 level or better, except between Unmasked and Covered ($p > .05$).

** Unmasked is different from Surrounded at .05 level, and from Covered at .001 level. Covered not different ($p > .05$) from Surrounded or Surrounded and covered.

3. In both the communication density and the speaker separation experi-

TABLE 2

DISTRIBUTION OF ATTENTION EFFECTS

Sources of Cont. Communication	Omissions with Attention to:		Mishearings with Attention to:	
	1 Speaker	2 Speakers	1 Speaker	2 Speakers
0 Density	.4%	.4%	.2%	.2%
1 Density	2.1%	3.0%	1.7%	3.7%
2 Density Exper.	4.7%	12.7%*	6.1%	3.3%
2 Separation Exper.	7.7%	14.6%**	4.9%	5.0%

* One Speaker-Two Speaker $p = .01$. ** One Speaker-Two Speaker $p = .001$.

ments, listening to the two speakers was compared with listening to only one of them. The results are given in Table 2 according to the number of sources of continuous communication.

4. In the directed attention experiment *S* was told when to listen for some of the most difficult calls of the two previous experiments, as well as the speaker to which to listen. Under these conditions the percentage of omissions was 1.2, as compared with 12.5% for the same calls in the speaker separation experiment when only one speaker had to be monitored, but the times of calls were not known, a significant difference. (The few remaining omissions were due to *S*'s listening to the wrong speaker when the two were placed one above the other.) Mishearings did not alter significantly, the percentages being 9.0 and 10.9, respectively.

5. In the speaker separation experiment an approximately 180° right-left horizontal angular separation was compared with a 3½–6° mid-line vertical separation. The effect upon omissions and mishearings is shown in Table 3, when only one and when both speakers had to be monitored.

The advantages of speaker separation were more marked with a high density of communication. In two additional experiments using no channel and one channel of continuous communication, the separation difference for omissions and mishearings combined was in neither case significant (mean of 12 and 16 *S*s, respectively).

6. In two quarters of the speaker separation experiment the block of conversations on one channel concerned *S*'s tower. (Where he had to listen only to one channel, it was always this channel which was involved.) These conversations were *similar* to the calls from visiting aircraft for which *S* had to listen, in that

TABLE 3

SPEAKER SEPARATION EFFECTS

Speakers Monitored	Omissions		Mishearings	
	Speakers Apart	Speakers Together	Speakers Apart	Speakers Together
1	4.9%	10.5%**	3.5%	6.3%*
2	13.2%	15.9%	3.5%	6.5%*

* Apart-Together $p = .05$. ** Apart-Together $p = .001$.

TABLE 4

EFFECTS OF SIMILARITY OF CONTENTS OF COMMUNICATIONS AS A FUNCTION OF NUMBER OF SPEAKERS MONITORED

Speakers Monitored	Omissions		Mishearings	
	Dissimilar	Similar	Dissimilar	Similar
1	6.2%	9.2%*	6.3%	3.6%
2	10.3%	18.9%**	5.1%	5.0%

* Dissimilar-Similar $p = .05$ (one-tail). ** Dissimilar-Similar $p = .01$.

both included the tower name. They could be distinguished in two ways: (*a*) Information was always requested and given in conversations, but never in calls from visitors; (*b*) *S* could soon become familiar with the numbers of the three aircraft holding the conversations with his tower in each of these two quarters of the experiment. In the other two quarters the conversations were *dissimilar*, in that they did not concern *S*'s tower. The effect of similarity upon the recorded calls from visitors is shown in Table 4 according to the number of speakers to be monitored.

In the communication density experiment the effect of similarity upon omissions was smaller when continuous communication was presented over only one channel, than when it was presented over both. However neither of the differences was significant.

The effect of similarity upon false selections (calls reported when no call was made to *S*) is shown in Table 5 for the speaker separation experiment, according to the arrangement of speakers. False selections have been expressed as a percentage of the actual number of calls made to *S*. When the two speakers were arranged one above the other, a tower name on the similar communication channel was quite often followed by a three-digit number on the other channel, and this combination was sometimes mistaken for a call from a visiting aircraft. At least one-fifth of the false selections were clearly caused in this way.

TABLE 5

EFFECTS OF SIMILARITY OF CONTENTS OF COMMUNICATIONS AS A FUNCTION OF THE ARRANGEMENT OF SPEAKERS

Speaker Arrangement	False Selections	
	Dissimilar	Similar
Apart	.9%	2.9%*
Together	1.6%	6.2%**

* Dissimilar-Similar $p = .05$.
** Dissimilar-Similar $p = .001$.

DISCUSSION

Omissions and inattention.—The following evidence suggests that omissions were associated with inattention: (*a*) The *S* was seldom aware when he had omitted a call (Result 1). (*b*) Surrounded calls (Table 1) were camouflaged between more or less similar communication on the same channel, and were therefore less likely to be noticed than unmasked or covered calls which came from an otherwise quiet loud-speaker, and therefore had an aspect of novelty (2). (*c*) Surrounded and covered calls (Table 1) were not only camouflaged in this way, but also had a competing source of continuous communication on the other channel to attract and hold *S*'s attention, and were therefore even less

likely to be noticed. (*d*) Calls were less likely to be noticed in a task involving listening to one of two sources of continuous communication when their times of occurrence were not known, than when *S* was told when to listen (Result 4). (*e*) They were still less likely to be noticed when the task involved listening to both the sources, and their times of occurrence were not known (Table 2). (*f*) Placing the loud-speakers together (Table 3) removed the directional cue, and thus left only the content of the communications as a means of distinguishing between the two sources, for the voices were the same. It therefore reduced the likelihood of calls from the correct source being noticed when only one source had to be monitored. (*g*) Increasing the similarity of the irrelevant material to the relevant (Table 4) reduced the likelihood of the relevant material being noticed, as with visual material (6).

Mishearings and physical interference.—In contrast to omissions, which were influenced by a number of attentional factors, mishearings depended upon the degree of physical interference between the two sources. Thus covered, and surrounded and covered, calls gave about the same number of mishearings, while both gave more than unmasked calls (Table 1). And placing the two speakers together increased the number of mishearings (Table 3), as previously reported in experiments using one speech source and one noise source (3, 4).

However, the just significant increase in mishearings of surrounded as compared with unmasked calls (Table 1), suggests that inattention may have been a minor determinant of mishearings. It cannot have been a very important one, as omissions and mishearings were distributed independently of each other (Result 1), and none of the other attentional factors significantly affected mishearings.

Attention in masking and speaker separation.—If the amount of correctly received relevant information is accepted as the criterion of efficiency, Table 1 shows that the surrounding type of masking can cause as great a loss of efficiency as covering. Surrounding and covering combined can cause a still greater loss. Previous work on speech masking (5) has been concerned only with covering and physical interference, for all the information from the source to which *S* was listening was relevant. Surrounding with its attentional effect can only occur by definition when relevant and irrelevant information are presented from the same source.

Table 3 shows that separating the two speakers not only reduced mishearings, by helping *S* to distinguish the composite noise from the two sources; it also reduced omissions when only one speaker had to be monitored, by giving him a clear indication of which speaker this was. The former effect has already been studied by Hirsh (3) and Kock (4), the latter by Broadbent (1). In the present experiment the attentional factor was at least as important by the efficiency criterion as the physical interference factor.

Summary

Selecting information from two simultaneous sources of speech was evaluated in terms of omissions (of which *S* was generally unaware) and mishearings, which were distributed independently. When relevant information was surrounded by irrelevant on the same channel, omissions exceeded mishearings. When it was covered by simultaneous irrelevant information on the other channel, mishearings predominated.

Instructions to listen to both sources (distributed attention) gave more omissions than instructions to listen to one of them (restricted attention). This in turn gave more than instructions as to when and to what to listen (directed attention). Mishearings were independent of these instructions. Placing the speakers together increased both omissions and mishearings. Similarity of relevant material to irrelevant increased omissions; it also increased false selections, especially with the speakers together. When only one or neither source presented information continuously, omissions and mishearings were reduced, and none of these differences was significant.

The results were discussed in terms of inattention and physical interference.

References

1. Broadbent, D. E. The role of auditory localization in attention and memory span. Cambridge. Appl. Psychol. Unit, Report No. 171, 1952.
2. Broadbent, D. E. Classical conditioning and human watch-keeping. *Psychol. Rev.*, in press.
3. Hirsh, I. J. The relation between localization and intelligibility. *J. acoust. Soc. Amer.*, 1950, **22**, 196–200.
4. Kock, W. E. Binaural localization and masking. *J. acoust. Soc. Amer.*, 1950, **22**, 801–804.
5. Miller, G. A. The masking of speech. *Psychol. Bull.*, 1947, **44**, 105–129.
6. Sleight, R. B., & Mowbray, G. H. Discriminability between geometric figures under complex conditions. *J. Psychol.*, 1951, **31**, 121–127.

(Received December 15, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

A RETEST FOR CONDITIONED INHIBITION IN THE ALPHABET-PRINTING TASK

RAYMOND E. SCHUCKER, LUCIA B. STEVENS, AND DOUGLAS S. ELLIS
Iowa State College

The present study was stimulated by Kimble's (3) and Wasserman's (7) experimental tests of a two-factor theory of inhibition (3). It investigates the possibility that these experiments overestimated the amount of conditioned inhibition in the alphabet-printing task.

The essential notion of the two-factor theory is that the inhibitory potential generated during relatively massed practice may be differentiated into two components, reactive inhibition (I_R) and conditioned inhibition ($_sI_R$). While both factors are assumed to be detrimental to performance, they are assigned somewhat different properties. I_R is thought to be a negative drive which arises as a result of response evocation. Since it is given the special property of spontaneously dissipating during rest, the attainment of rest becomes a rewarding or reinforcing state of affairs for the organism. $_sI_R$ is assigned the status of a habit which develops as a result of the resting response becoming conditioned to stimuli present in the situation. As a habit, or learned resting response, $_sI_R$ does not dissipate spontaneously and thus may be permanently detrimental to performance. The amount of I_R generated by a sequence of responses can be indexed, according to Kimble, by the amount of reminiscence shown after introduction of a rest of proper length. The difference between postrest performance and performance obtained under highly distributed practice is suggested as a measure of $_sI_R$.

The experimental findings which lend the most support to the two-factor hypothesis are those of Kimble (3) and Wasserman (7). Kimble used a reversed alphabet-printing task and found highly significant amounts of both I_R and $_sI_R$ in massed-practice (MP) groups. MP groups which received an interpolated rest showed a significant increase in postrest performance (reminiscence, the I_R index) and their postrest performance was significantly lower than the performance of a distributed-practice (DP) group (the difference between the two groups is the $_sI_R$ index). Wasserman elaborated Kimble's study by repeating it at conditions of high and low motivation and found essentially the same result.

However, analysis of Kimble's experimental procedure suggests that certain uncontrolled variables could explain the large amounts of $_sI_R$ found. First, Kimble does not mention what the groups did during rest periods. Assuming that no rest-interval activity was introduced, it is quite possible that rehearsal took place both in the intertrial rests of the DP group and the single interpolated rest of the MP groups. If it is assumed that distributed rehearsal is more effective than massed rehearsal, rehearsal would be more beneficial to DP than to MP groups. This would, of course, inflate estimates of $_sI_R$. This criticism also applies to Wasserman's study in which *S*s were asked to sit quietly during rests.

Second, it is not clear whether or not knowledge of results was given to *S*s. If no systematic knowledge was given, *S*s in the DP group could have used their unfilled intertrial rests to

obtain knowledge of results. No such opportunity would have been available to MP *Ss*. This motivational difference would operate to separate further the performance curves of the two groups, and produce spuriously high estimates of ${}_{S}I_{R}$. It appears that Wasserman's study is open to the same criticism.

A third uncontrolled variable in Kimble's study was corrected by Wasserman. Kimble did not equate his DP and MP groups for total number of responses evoked. Instead, the groups were compared after the same number of trials. This procedure penalizes the MP groups, since they have had less opportunity to learn (had fewer response evocations) than the DP group. This factor, like the others, would inflate estimates of ${}_{S}I_{R}$. Wasserman corrected the data of his study for this practice differential; the correction reduced measured ${}_{S}I_{R}$, but did not eliminate it.

The present experiment is designed to test the hypothesis that the ${}_{S}I_{R}$ found by Kimble and by Wasserman is due to the factors listed above. The experiment repeats portions of Kimble's study using the following controls for these factors: (*a*) Colored photographic slides were used as a rest-filling activity; (*b*) test sheets were constructed to provide a designated space for each letter printed, and these spaces were numbered to provide knowledge of results.

A final feature of the experiment was dictated by previous results and the rationale of the problem. In an unpublished study conducted by members of the experimental psychology course at Iowa State College during the winter quarter of 1952, it was found that these controls did prevent the appearance of measured ${}_{S}I_{R}$, i.e., MP groups, after rest, did recover to the level of DP groups. Thus, in the present experiment, small postrest differences between MP and DP groups are expected. In addition, we are in the disadvantageous position of attempting to provide evidence for a null hypothesis. Both considerations dictate obtaining the most precise data possible. In the present experiment an attempt was made to provide this precision by introducing practice trials, and using these data as a covariate in the tests of significance.

The introduction of practice trials raises the additional problem of what distribution condition should be used in these trials. For example, if MP is used, then *Ss* who subsequently shift to DP may be penalized by the failure of complete learning transfer to occur. Or they may be benefited by the failure of inhibition to transfer completely. An analogous argument can be raised for the case where DP practice trials are used. In the present study we have not overcome this difficulty, but have provided a check on it. MP practice trials have been used for all groups except a special group. This special group receives DP throughout practice and experimental trials. Comparison of its performance with the DP group which receives MP practice trials permits a check on any differential transfer experienced by the latter group.

Method

Subjects.—The 118 *Ss* who furnished usable data (8 *Ss* had to be dropped because they did not follow instructions or were confused by the task) were male and female students in the introductory psychology course at Iowa State College during the spring quarter of 1952. They served in the experiment as part of the requirement of that course. Several weeks before the experiment, *Ss* were given an opportunity to choose a desirable time to serve in the experiment and, on the basis of the number of requests for a given hour, the experimental groups were formed. These groups were then randomly assigned to experimental conditions.

Experimental conditions.—Kimble's procedure for measuring ${}_{S}I_{R}$ requires two basic experimental conditions. Condition I, a MP condition, received an interpolated rest after a given amount of practice. Condition II is a DP condition, against which the postrest performance of Cond. I may be compared. Kimble used six Cond. I groups, each representing a specified amount of prerest practice. Since his obtained ${}_{S}I_{R}$ was greatest for the Cond. I group, which received the most prerest practice (30 trials), the present study used only two Cond. I groups. Groups MP-20 and MP-30 of the present study received a rest after 20 and 30 MP trials, respectively.

Our Cond. II group, Group MP–DP, received the 30-sec.-work–30-sec.-rest degree of distribution used by both Kimble and Wasserman. The fourth group of the present study, Group DP–DP, differs from the other groups in that it received DP practice trials. Since in all other respects it received the same treatment as Group MP–DP, it provides a check on any transfer effects generated in Group MP–DP by the MP practice trials.

The *N*'s of these various groups were: Group MP-20, 30 (all males); Group MP-30, 30 (29 males, 1 female); Group MP–DP, 28 (all males); Group DP–DP, 28 (27 males, 1 female).

Procedure.—At the start of a session, *S*s were informed that the experiment was concerned with the effects of work on their ability to memorize slides. They were told that both the alphabet printing and the slides were important parts of the experiment. In regard to the slides, *S*s were told that they would be given a test covering the slide material at the end of the experiment. Such a test, consisting of ten true-false and eight completion questions, was actually administered at the end of the session. In regard to the printing task, *S*s were given careful instructions and a demonstration of the method to be used. At each new trial they were to start with the next letter of the alphabet, were to print as fast as possible, and were not to bother to correct mistakes once they had been made. Two pencils were provided for each *S* to minimize breakage delays. Signals to change trials were given both verbally and by a bell signal.

As in the Kimble and Wasserman studies, 30-sec. trials were used throughout the experiment. All groups received five practice trials. No intertrial rest separated the first and second practice trials, but the groups were given a chance to ask questions and check their work after the second practice trial. Subsequently, Groups MP–20, MP–30, and MP–DP received three MP practice trials, which served as a covariate in the later analysis of the data. Group DP–DP also received three additional practice trials, but these trials were separated by 30-sec. intertrial rests.

Following the practice trials, all groups were given a 5-min. rest. The groups then performed 33 trials under the following conditions: Groups MP-20 and MP-30: MP before and after a 10-min. rest inserted after Trial 20 or 30; Groups MP–DP and DP–DP: continuous DP.

Apparatus.—The sheets for the reversed alphabet-printing task contained ten trials, with the trial spaces subdivided into 60 equal $\frac{1}{2} \times \frac{1}{2}$-in. spaces. Within a trial, each of these spaces was consecutively numbered so that *S*s would have knowledge of results at all times. Each *S* received four of these sheets stapled together. Since spaces for the five practice trials were on the first sheet, *S* had to turn pages after Trials 5, 15, and 25.

During all rest periods, both in practice and experimental trials, colored slides of campus activities of general interest to the students were shown. In order that the slides be shown clearly, it was necessary to turn off the room lights when they were shown. However, there was enough illumination in the room at all times for *S*s to see their printing sheets. A Model 2*a* Kodaslide projector equipped with a 100-w. bulb was used, and slides were shown on a 4 × 5-ft. screen.

Results

The basic means of the experiment are plotted in Fig. 1. Following Kimble's and Wasserman's procedure, no correction has been made for errors in printing. The performance dips at Trials 6, 16, and 26 can be attributed to *S*s' having to turn pages of the printing task at these points. The critical consideration for the present study is whether the postrest performance of the MP groups recovers to the performance of the DP groups (indicating no ${}_{S}I_{R}$). Such recovery appears to have occurred in Group MP-30, but is not so evident in Group MP-20. The statistical analysis of these recovery effects will be given after two other aspects of the data have been considered: (*a*) the initial ability of the groups; (*b*) the differences between the two DP groups.

Initial ability of groups.—The four groups show very similar performance on practice trials iii, iv, and v. Three of the groups, MP-20, MP-30, and MP–DP, were treated alike (received MP) during these practice trials and thus may be compared on initial ability. Analysis of variance of these groups' scores summed over the three practice trials yields an F of 1.21, which is not significant at the 5% level of confidence. Since the fourth group, DP–DP, received DP practice trials, analysis of its practice-trial data would not yield information regarding its relative initial ability.

Comparisons of DP groups.—Comparison for the two DP groups provides a check on whether MP practice trials have penalized Group MP–DP. Although MP–DP and DP–DP attain comparable performance levels in the latter stages of practice, there is some indication that MP–DP is relatively depressed during Trials 2 and 3. However, analysis of variance indicates not only that this temporary

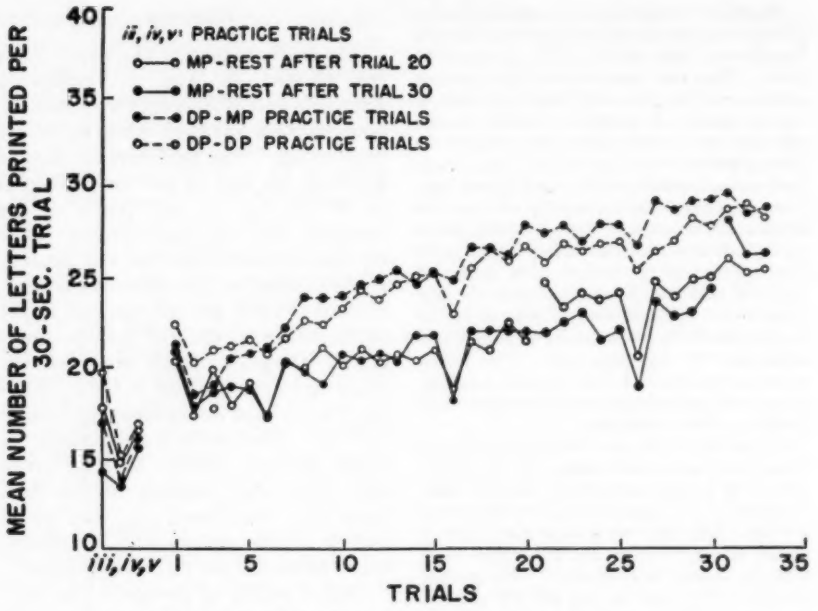


FIG. 1. Mean letter-printing performance of the four groups

depression is not statistically significant, but also that the two DP groups do not differ significantly in total letters printed during the 33 experimental trials. We conclude that MP–DP provides just as good an estimate of DP performance as DP–DP, with the advantage of having practice-trial data suitable for use in subsequent analyses of covariance.

Statistical significance of sI_R index for MP groups.—As indicated previously, Wasserman's correction must be used in this connection. The correction involves choosing an appropriate point on the DP curve so that comparisons of DP performance with postrest MP performance are made at points where the groups have had an equal opportunity to learn the printing habit (have printed an equal number of letters). Calculation indicated that at Trials 18 and 26 Group MP–DP had printed the same total number of letters as Groups MP-20 and MP-30 had achieved by Trials 20 and 30, respectively. In the critical postrest comparison we then use the next trial for each group, i.e., Trials 19 and 27 for Group MP–DP and Trials 21 and 31 for Groups MP-20 and MP-30, respectively.

The relevant analyses of covariance, which test whether initial postrest performance of an MP group is comparable to the performance of Group MP–DP at the appropriate trial, are arrayed in Table 1. The table indicates that the null hypothesis, which is identical with the hypothesis of no sI_R, cannot be rejected at the 5% level of confidence. The covariate (letters printed on practice Trials iii, iv, and v) is singularly helpful in reducing the error term for these tests. The variate-covariate correlations for Table 1 are, for Groups MP-20 and MP-30, respectively, .697 and .640,

TABLE 1

COMPARISON BY ANALYSIS OF COVARIANCE OF POSTREST MP PERFORMANCE WITH DP PERFORMANCE

Comparison	Mean Letters Printed	Source of Variance	*df*	*MS*	*F**
MP–20 at Trial 21	24.8	Between	1	43.0	3.19
vs.		Within	55	13.5	
MP–DP at Trial 19	26.3	Total	56		
MP–30 at Trial 31	27.8	Between	1	64.2	3.74
vs.		Within	57	17.1	
MP–DP at Trial 27	28.8	Total	58		

* For 1 and 55 *df*, *F* is 4.02 for $p = 5\%$. For 1 and 57 *df*, *F* is 4.01 for $p = 5\%$.

which correspond, respectively, to approximately a 50% and a 40% reduction in error variance.

DISCUSSION

As indicated by Table 1, the data of the present experiment are consistent with the hypothesis that the introduction of certain experimental controls would eliminate measured ${}_{S}I_{R}$ in the alphabet-printing task. The fact that statistical tests utilizing a very relevant covariate fail to reach significance indicates that, for our data, the hypothesis has substantial probability.

Since the present study was designed for the purpose of controlling factors thought to have influenced past experimental results, it is anticipated that a number of procedural questions may arise. One immediate difficulty is that, although the performance of our MP *S*s is comparable to that of Kimble's MP *S*s, our DP *S*s, once they have passed Trial 10, print 4–6 fewer letters per trial than Kimble's DP *S*s. This is in spite of the fact that our DP *S*s have the advantage over Kimble's of receiving both practice trials and regular knowledge of results. It is this poorer performance of our DP *S*s which eliminates measured ${}_{S}I_{R}$. Their poorer performance might be due to one of three factors: (*a*) the slide-watching activity may have prevented the rehearsal which seemed probable in Kimble's study; (*b*) the slide-watching activity may have generated habits inconsistent with printing habits; (*c*) the instructions to *S* may have directed his efforts primarily towards slide watching at the expense of alphabet printing. If the results are due to (*a*), no procedural question arises—the rest-filling activity has functioned as desired. Factor (*b*) seems unlikely. It is difficult to locate similarities between alphabet printing and slide watching which could mediate habit interference. Factor (*c*) is a possibility. Whether or not it is potent in the present study cannot be determined by the data at hand. A study in which instructions to *S* are varied in respect to the relative emphasis placed upon the printing and slide-watching tasks should answer the question.

A second question concerning experimental procedure has to do with the control of illumination. It might be argued that the performance of the DP groups was detrimentally affected by the turning on and off of the room lights, the point being that light adaptation would have to take place in going from the lower illumination which accompanied slide viewing to the high illumination (room lights on) when printing the alphabet. Although some adaptation did have to take place, the difference in the two conditions of illumination was not great. The *S*s were able to see their work sheets at all times.

A third procedural question has to do with the effects of the initial practice trials. Group MP–DP gives no evidence of having been permanently penalized by its MP practice trials. However, there is the possibility that the practice trials developed inhibition which would later penalize the MP groups in comparison to the DP groups. Several arguments may be advanced against considering this possibility seriously. First, a rest intervened between the second and third practice trials. Second, a 5-min. rest was inserted after the practice trials to dissipate any small amount of inhibition which might have been generated. Even if we assume that this rest was not sufficient to dissipate practice-trial I_R, this factor would not be critical in the present experiment. The DP groups could dissipate any remaining I_R during their intertrial rests, while the MP groups would not have this opportunity. Thus, any carry-over of I_R from the practice trials would tend to depress MP performance

and inflate measures of ${}_sI_R$. Third, although the 5-min. rest would not dissipate any ${}_sI_R$ generated during practice periods, Kimble's data (3) show no appreciable ${}_sI_R$ even after five massed trials. These arguments lead us to believe that the findings of the present experiment cannot be attributed to inhibition developed during the practice trials.

In general, we believe that these procedural questions do not invalidate the present finding of no ${}_sI_R$ in the alphabet-printing task. We are inclined toward the conclusion that, for this task, ${}_sI_R$ is an unnecessary construct for describing performance differences between MP and DP. However, on the basis of present evidence, the generality of this conclusion seems problematical. The present experimental design, like that used by Kimble and by Wasserman, is based on several unproven assumptions: One assumption is that a 10-min. rest is optimal for the generation of maximum reminiscence in this task. A second is that the 30-sec. work–30-sec. rest DP condition is optimal. A third is that no reminiscence would be shown in this DP condition if it were subjected to a 10-min. interpolated rest.

In addition, there is the problem of generalizing the results to other motor tasks. Although some studies (**1, 4, 6**) with other motor tasks, such as the pursuit rotor, find only fragmentary evidence for ${}_sI_R$, recent papers by Barch (**2**) and Kimble and Shatel (**5**) indicate substantial postrest differences between MP and DP. Whether the introduction of controls of rest-interval activity and knowledge of results would remove such indications of ${}_sI_R$ in these pursuit tasks remains a problem for further study.

Summary

An experiment was conducted which incorporated certain procedural controls (specific activity for all rest intervals, systematic knowledge of results) in testing for conditioned inhibition in the alphabet-printing task. Two distributed-practice groups and two massed-practice groups with a total N of 118 were employed. The distributed-practice groups alternately printed letters and looked at photographic slides. The massed-practice groups continuously printed letters except for a 10-min. rest, which was filled with slide viewing, following either Trial 20 or Trial 30. All groups completed a total of 33 trials. The hypothesis tested was that the measured conditioned inhibition found in previous studies using the same task was due to failure to control adequately rest-interval activity and knowledge of results. The experimental results supported the hypothesis.

References

1. Ammons, R. B. Acquisition of motor skill: III. Effects of initially distributed practice on rotary pursuit performance. *J. exp. Psychol.*, 1950, **40**, 777–787.
2. Barch, A. M. The relative contributions of I_R and ${}_sI_R$ to the total inhibitory potential. *Perc. & mot. Skills Res. Exch.*, 1952, **4**, 27–31.
3. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
4. Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
5. Kimble, G. A., & Shatel, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
6. Reynolds, B., & Bilodeau, I. McD. Acquisition and retention of three psychomotor tasks as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, **44**, 19–26.
7. Wasserman, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

(Received November 28, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

PATTERNING EFFECT IN PARTIAL REINFORCEMENT

JOSEPH H. GROSSLIGHT, JOHN F. HALL, AND JOSEPH MURNIN
The Pennsylvania State College

Early stimulus-response learning theorists had difficulty in accounting for the greater resistance to extinction among those organisms which were provided with partial reinforcement than among those provided with continuous reinforcement. Jenkins and Stanley (**4**) in a recent article have summarized the experiments and theorizing on this problem.

Recently Sheffield (**5**) has postulated that the greater extinction resistance of partial reinforcement is based upon *the aftereffects*[1] *of nonreinforcement in the original training situation.* The authors would like to offer a related aftereffects hypothesis, making more explicit, however, one fundamental relationship not previously stated as such, namely, the patterning effect in the partial-reinforcement situation.

Grosslight and Child (**2**) pointed out that the occurrence of nonreward in a series of trials permits *the response of persisting after failure* to occur and be rewarded. In the typical aperiodic partial-reinforcement situation, the occurrence of a reinforced response immediately following a nonreinforced response (RUR) could result in this learning to "persist after failure." On the other hand, the occurrence of a reinforced trial followed by a nonreinforced trial (provided that no further reinforcement occurs in the series) will not provide this differential learning.

In order to assess this patterning effect in partial reinforcement, it is necessary to use a partial-reinforcement procedure for one group in which the nonreinforcement–reinforcement (UR) sequence occurs, and for another in which it does not occur. In the latter case the association of the cues of nonreinforcement with reinforcement is avoided. Under these conditions, it is predicted that the extinction rate of the reinforcement–nonreinforcement (RU) group should be faster than the nonreinforcement–reinforcement (UR) group. The reasoning behind this deduction can be made clearer when it is remembered that the basic thesis concerns the nonreinforcement–reinforcement associational aftereffects for one group during training and the absence in the other. These aftereffects will operate in extinction to produce an increased response frequency.

It is also predicted that the UR group will manifest a longer extinction time than the continuously reinforced (RR) group. A differential prediction as to extinction rate between the RU and RR groups is less clear-cut. In terms of the absence of the patterning effect and the fewer reinforcements for the RU group, it could be postulated that this group (RU) would extinguish at a faster rate than the continuous (RR) group. However, the possible operation of other variables emphasized by a stimulus-generalization point of view would make such a prediction less probable. In light of the procedure and the nature of the *S*s in this experiment, no differential test of the above is possible.

Design and Procedure

The *S*s were 192 college students of both sexes, obtained from the elementary psychology classes at The Pennsylvania State College. No attempt was made to assign *S*s to the respective experimental conditions individually; rather, 32 *S*s were first randomly selected from each class, and then each block of 32 was randomly assigned to one of the six experimental groups. Twenty

[1] Sheffield's aftereffect explanation is in terms of a stimulus-generalization hypothesis. It is equally possible, however, to have the aftereffects related to a secondary-reinforcement hypothesis.

TABLE 1

PATTERNING OF TRAINING TRIALS BY SERIES FOR THE THREE EXPERIMENTAL CONDITIONS

Exp. Cond.			Exp. Cond.			Terminal Training and Extinction All Groups
RR	RU	UR	RR	RU	UR	
Series 1			Series 4			R
R	R	R	R	R	R	R
R	R	U	R	R	R	U
R	U	U	R	R	R	U
R	U	R	R	R	R	U
Series 2			R	R	R	U
R	R	R	R	R	R	U
R	R	R	R	R	U	U
R	R	R	R	U	U	U
R	R	R	R	U	U	U
R	R	U	R	U	R	U
R	U	U	Series 5			U
R	U	U	R	R	R	
R	U	R	R	R	U	
Series 3			R	U	U	
R	R	R	R	U	R	
R	R	R				
R	R	U				
R	U	U				
R	U	R				

Ss were discarded because their records indicated failure to follow instruction; the final *N* was thus 172.

A modification of Humphreys' (3) original verbal-expectancy procedure was employed. The availability of the Classroom Communicator of the Instructional Film Research Program[2] afforded an excellent opportunity for the presentation of two lights in a fixed pattern as well as for rigid control and recording of *Ss*' responses. The *Ss* were required to indicate whether or not they expected a second light to follow the first light. Each *S* indicated his "expectancy" by pressing one of two levers on a private control panel attached to his chair. Choices were automatically recorded.

All training was accomplished in a single session requiring approximately 40 min. The total experimental session was divided into five training and one extinction periods. The training periods consisted of a total of 33 trials, followed by an extinction period of 10 trials. In order to prevent any set resulting from each period having a fixed number of trials, the number of trials per training period was varied. The pattern of trials and number of reinforcements for each group per period can be ascertained from Table 1. One point of procedure should be emphasized, because it may have resulted in some confounding of the pattern variable as tested in this study. It is to be noted (Table 1) that nonreinforcement (U) follows nonreinforcement (U) for both experimental groups. It was felt that a single nonreinforced trial might not be sufficient to demonstrate the operation of the UR patterning variable.

One theoretical question with regard to the concepts of reinforcement and extinction was a concern. Detambel (1) has pointed out that Humphreys' procedure does not follow the usual extinction paradigm: differential responses are set up during training, but only one is measured during extinction. In an attempt to test the importance of this variable, three of the six groups were given the experimental treatments (UR, RU, and RR) with a monetary reward based on the highest scores as determined by a point system. This point system (see instructions to *Ss*) would seem to result in only one response being clearly rewarded, namely, a correct expectation of the light going on. However, this extinction-procedure variation as recommended by Detambel would not seem to alter the basic hypothesis of this study, namely, the patterning effect as reflected between the RU and UR groups.

The 32 *Ss* for each experimental group were trained and tested in a single experimental session. The following instructions were read to them: "You are about to participate in an experiment, so please follow the instructions carefully and give us your full cooperation. In the front of the room you see a box containing two lights. On every trial the light on your left will go on; the light on your right may or may not go on. When the light on your left goes on, you are to decide whether you expect or do not expect the

[2] The Classroom Communicator is an instrument designed for simultaneous recording of responses for a large number of *Ss*. This instrument was developed by the Instructional Film Research Program, a research organization sponsored by the U. S. Navy.

TABLE 2

MEAN NUMBER OF LIGHT EXPECTANCIES DURING THE EXTINCTION SERIES

Condition	Monetary			Nonmonetary			Combined		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
UR	30	5.66	2.76	27	5.52	2.54	57	5.60	2.65
RU	28	4.04	2.74	29	3.28	2.57	57	3.65	2.68
RR	29	2.10	.96	29	1.90	.66	58	2.00	.84

light on your right to go on. You will indicate whether you expect or do not expect the light on your right to go on in the following manner: on each chair arm there is a black box containing five keys. (The operation of the keys for choice indication was explained.) Indicate your choice and do not change your answer. A new trial begins each time the light on your left goes on. The trials will come in several series. At the end of each series the *E* at the front of the room will tell you that the end of the series has occurred. He will tell you when the new series begins." (The *S*s were then asked for questions and given a demonstration trial.)

Those given monetary reward were instructed further as follows: "You will be scored on your ability to anticipate correctly whether the light on the right will go on or not go on. To the person who makes the highest score we will give a prize of two dollars, to the next highest person a prize of one dollar. The score in this experiment will be determined in the following manner: If you indicate that you expect the light on the right to go on, and it does go on, you will be given two points. If you indicate that you expect the light on the right to go on, and it does not, you lose one point. If you indicate that you do not expect the light on the right to go on, and it does, you will lose a point. If you indicate that you do not expect the light on the right to go on, and it does not, you will neither gain nor lose a point for that trial."

Following the above instructions, the *E* operating the Communicator informed *S*s as to the series presented, i.e., Series 1, etc. The instrument was set to present the trials at specific time intervals: the first light on the left came on and was followed in 3 sec. by a chime at which time *S* made his choice; 2 sec. later all response boxes were locked and the light on the right appeared or did not appear according to the particular trial patterning. At the end of the series, *E* announced: "That was the end of series Number 1. This is the beginning of series Number 2," etc.

RESULTS AND DISCUSSION

The analyses of the extinction series for both the monetary and nonmonetary conditions indicated that the UR group had the greatest average light expectancy, with the RR group showing the smallest (Table 2). *t* tests computed between similar experimental treatments for the monetary and nonmonetary conditions revealed no significant differences. The separate *t* tests were necessitated as Bartlett's test indicated significant nonhomogeneity of variance.

The problem of differential extinction procedure with reference to the Detambel criticism and as interpreted by this paper did not lead to relevant extinction differences. Therefore, more reliable data for the varying experimental treatments were obtained by combining the data from the monetary and nonmonetary conditions (Table 2).

The average extinction light expectancy for the UR group is significantly larger than that for either the RU group ($t = 3.86$, $p > .01$) or for the RR group ($t = 9.67$, $p > .01$). These comparisons between the UR group and the other groups when tested separately for the monetary and nonmonetary conditions are in the same direction and significant at the .05 level of confidence or better.

The trial-by-trial analysis during extinction for the three experimental treatments (Fig. 1) makes the pattern-

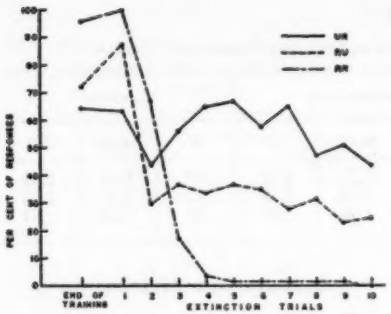


FIG. 1. Percentage of light expectancy per trial for the experimental groups during extinction

ing effect even more apparent. It is relevant to note that trial-by-trial analyses for the UR and RU groups *during training* show no basic differences. These results would seem clearly to support the major patterning hypothesis of this paper. Apparently this patterning effect operated in the apparent absence of *S*'s awareness. The *S*s interrogated at the end of the experiment were unable to verbalize any consistent light order, i.e., patterning, running through the training sessions.

A differential prediction of resistance to extinction for the RU and RR groups was not clearly set forth. This experiment demonstrated that in all comparisons, the RU group had a significantly greater resistance to extinction as shown by a larger average light expectancy than did the RR group (Table 2). This difference for the combined data results in a t of 4.39, which is significant at the .01 level of confidence. At least two possible explanations of this finding can be offered. First, the use of human *S*s and the consecutive training series which always began with reinforcement of the second light may have resulted in verbal mediation. In this fashion, some reinforcement would follow nonreinforcement even for the RU group. Secondly, from a stimulus-generalization point of view, there is greater "similarity" between acquisition and extinction for the RU group than for the RR group.

SUMMARY

The purpose of this experiment was to demonstrate the operation of a patterning variable, i.e., reinforcement–nonreinforcement sequence, in partial reinforcement. A modification of the Humphreys light-expectancy procedure was employed with 172 college students. Three conditions consisted of continuous reinforcement (RR), partial reinforcement with nonreinforcement termination (RU), and partial reinforcement with reinforcement termination (UR). The results indicated that nonreinforcement followed by reinforcement leads to the greatest resistance to extinction. The general theoretical importance and difficulties of patterning in partial reinforcement were discussed.

REFERENCES

1. DETAMBEL, M. H. A reanalysis of Humphreys' verbal expectations. Unpublished M.A. thesis, Indiana Univer., 1950.
2. GROSSLIGHT, J. H., & CHILD, I. L. Persistence as a function of previous experience of failure followed by success. *Amer. J. Psychol.*, 1947, **60**, 378–387.
3. HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
4. JENKINS, W. O., & STANLEY, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
5. SHEFFIELD, V. R. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–525.
6. SNEDECOR, G. W. *Statistical methods.* Ames, Iowa: Iowa State Univer. Press, 1946.

(Received November 26, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

THE DEVELOPMENT OF NONCONTINUITY BEHAVIOR THROUGH CONTINUITY LEARNING[1]

LYNE STARLING REID
University of Virginia

Certain recently reported experiments suggest the possibility that the schism between the continuity and noncontinuity positions with respect to discrimination learning may be narrowing. Particularly the studies of Harlow (**1, 2**) on the development of "learning sets" and Lawrence (**3, 4**) on the acquired distinctiveness of cues would seem to indicate that the behavioral observations which led the noncontinuity group to posit the constructs of "hypotheses" and "attempted solutions" are capable of being better understood in terms of the more elemental principles developed by the continuity theorists. Within this framework there seem to be two directions that one may take in attempting to bridge the gap between the divergent theoretical interpretations. The first would be to substitute for "hypotheses" and "attempted solutions" the learning of some mediating response capable of being derived from the continuity principles. Responses such as these would be Harlow's "learning sets," the mediating mechanism suggested by Lawrence to account for the acquisition of cue distinctiveness, or the one to be suggested in the present paper. The second is a continued effort to describe the behavioral phenomena exclusively in terms of the lower-level concepts usually associated with continuity principles, e.g., the excitatory or inhibitory potential existing between a given stimulus and response, stimulus generalization, etc. (**6**). While at present the question remains open as to which of the two approaches is the more fruitful, the experiment to be reported may provide one further step toward the solution of the problem.

Independent of the two viewpoints it is reasonable to assume from the Harlow work that an organism learning a sequence of new problems is developing of necessity a complex habit structure. It is possible to conceptualize this structure as being made up in its entirety of many specific discrete S–R habits which interact in terms of the usually accepted continuity principles. On the other hand, and particularly in view of the Harlow results, it seems more likely that as learning progresses, there evolve habits which play new kinds of roles in the animal's behavior. It is suggested by the present paper that the animal is learning to discriminate a specific set of stimuli within the total stimulus complex. For example, in the Lashley discrimination apparatus the rat learns to jump to a specific stimulus card, but he also learns to look at both cards before jumping—a response that did not consistently take place when the animal was first placed in the problem situation.

What is being hypothesized is that as the rat learns to make a specific choice discrimination, he is also learning a response of discriminating, i.e., learning to respond to a set of stimuli of which the specific stimulus is a member. Since current continuity and noncontinuity theories both predict that the overlearning of a par-

[1] This work was supported in part by a research grant from the Research Committee of the University of Virginia. Much of the material was presented at the 1951 meeting of the American Psychological Association.

ticular reward relationship between two stimulus objects would result in a greater habit interference in the learning of the reversal of this relationship, evidence for the above hypothesis can be obtained by a demonstration of a more rapid learning of the reversal by a group that has received an excessive amount of overtraining to the original reward relationship, i.e., has had an opportunity to develop more fully a response of discriminating. The experiment to be described is such a test of the hypothesis.[2]

Procedure

Subjects.—Forty-five female albino rats of the Wistar strain, aged 90 to 120 days at the beginning of the experiment, were used in this investigation. The rats were from the animal colony maintained by the Psychological Laboratory of the University of Virginia.

Apparatus.—The apparatus consisted of a straight alley runway, and a Y-alley discrimination apparatus both of which were painted flat black.

The runway was 30 in. long with a starting box at one end and a goal box at the other. Its floor was 6 in. wide and its walls 6 in. high. The alley was covered with hardware cloth. The starting box was 6 in. square on all sides and had a hinged top through which *S*s could be introduced. A vertical sliding door located in the front of the starting box at the beginning of the alley runway was the same width and height as the alleyway. The goal box was 8 in. high, 10 in. wide, and 10 in. in length. An opening 6 in. wide and 5 in. high was on the front side of the box. On the inside of the box a small metal rod was placed above the opening upon which cards large enough to cover it could be hung. The box had a hinged top made of hardware cloth.

The discrimination apparatus was similar to a Y-alley maze. From the starting box to the bifurcation was a distance of 12 in. The arms of the Y were 14 in. in length and located at the end of each arm was a goal box. Dimensions of the floor and walls of the alleys and those of the starting box and goal boxes were the same as those for the runway described above.

The apparatus was lighted throughout the experiment by two shielded 40-w. white fluorescent tubes suspended longitudinally 52 in. above the table upon which it was placed. This provided an even distribution of light with no shadows cast by any part of the apparatus upon another.

Stimulus cards were made of medium-weight cardboard and were 7 in. wide and 5¼ in. in height. Two hooks were attached to the upper edge of each card so that it could be hung from the metal rods above the goal-box doors. They were so arranged that the card marking the correct door was allowed to swing freely, thus permitting the rat to gain entrance into the food box, while the card denoting the incorrect door could be made fast. Because the cards were wider than the doors, they could swing in only one direction, and once *S* had entered the box, it was impossible for it to get out and enter other parts of the apparatus. Three stimulus cards varying in brightness along a black-white continuum were used. The closest match on the Munsell color guide for these cards was N2/0 (black), N5/0 (gray), and N9/0 (white).

Preliminary training.—The *S*s were handled for seven days previous to the experiment proper to habituate them to the experimental procedure. On each of the last three of these days they were run two trials in the straight alley runway and permitted access to food in the goal box. The gray card was used to cover the opening through which they gained admission to the goal box. During these seven days and throughout the remainder of the experiment water was always in their home cages, but *S*s received food only during the experimental period and in special feeding cages for 1 hr. thereafter.

Discrimination training and overlearning.—With the black card as the positive stimulus each *S* was run five trials per day in the discrimination apparatus. These trials were so spaced that no trial followed another in less than 3 min. The cards were alternated with respect to position in a prearranged random fashion during each period of five trials with the only restriction that there were an equal number of reinforced responses to the right or left during a ten-day period. This series was repeated every ten days during both the discrimination, overlearning, and reversal training.

The self-corrective technique was utilized for all training in the discrimination apparatus; a response being considered an error if the rat touched the incorrect card. Thus, on each trial, there was reinforcement with respect to the positive stimulus card.

The *S*s continued to receive five trials a day until they reached a criterion of 9 out of 10 correct responses with the further requirement that the last five responses all had to be correct. Upon reaching the criterion they were divided into three groups in a counterbalanced order to

[2] A different interpretation has been offered by Spence. This will be discussed later.

provide matched groups with respect to habit strength in responding to the black stimulus card. Group CO, the control group, was changed immediately to the reversal training described below. Groups E50 and E150, respectively, were given 50 and 150 additional reinforced trials to the black stimulus cards. At the completion of this training they, too, were shifted to the reversal schedule.

Reversal training.—During this period of the experiment the positive stimulus card was the white one. As a result, the response now to be learned was the exact opposite of the response learned during the original discrimination training. The daily procedure for running *Ss* was the same as that which had been previously maintained, and the stimulus cards were again systematically varied according to the prearranged schedule. The training was continued until a criterion identical with the one outlined above was reached. In this period all three groups received the same training.[3]

Results

The means and *SD*'s of the number of trials to reach the criterion in the original discrimination learning are summarized for the three groups in Table 1. It can be seen by inspection that none of the mean differences between these groups is statistically significant.

The major concern of the investigation was to determine whether or not significant differences could be found among the three groups with respect to the number of trials necessary to reach the criterion in the discrimination reversal. Table 2 presents the means and *SD*'s of the groups during this period of their training for both the number of trials to reach the criterion and the number of erroneous intrusions in terms of the original discrimination habit.

[3] That the results to be described are not restricted to the specific conditions of this experiment has been demonstrated by subsequent research in the Virginia laboratory. The same results have been obtained when: (*a*) the apparatus was painted an intermediate gray; (*b*) the noncorrective technique was used; (*c*) both original and reversal learning were to both white and black stimulus cards.

TABLE 1

Trials to Criterion in Original Discrimination Learning
(N = 15 each group)

Group	Mean	*SD*
CO (0 trials past criterion)	80.3	17.5
E50 (50 trials past criterion)	87.7	18.0
E150 (150 trials past criterion)	78.3	16.0

TABLE 2

Trials to Criterion and Erroneous Intrusions in Discrimination Reversal Learning

Group	Trials		Intrusions	
	Mean	*SD*	Mean	*SD*
CO	138.3	29.8	2.0	3.0
E50	129.0	52.1	8.0	6.5
E150	70.0	26.5	16.0	10.2

An analysis of variance for matched groups when applied to the number of trials to reach the criterion yields, for the overlearning treatment, an F of 17.02, which is significant for 2 and 28 *df* at the 1% level of confidence or beyond. A comparison of the means shows that large differences exist between Groups E150 and CO (t = 6.21, .1% level); and between Group E150 and Group E50 (t = 4.87, .1% level); but that only a negligible difference exists between Groups E50 and CO (t = .63). Since the first two comparisons are well beyond the .001 probability level, it seems apparent that *Ss* carried 150 trials past the criterion in the original discrimination training learned the discrimination reversal in significantly fewer trials than did either the group with zero added trials or the group with only 50 added trials. Failure to find a significant difference between Groups CO and E50 was not entirely unexpected. With 50 trials past the criterion Group

E50 had perhaps developed close to a maximum habit strength for the original interfering response to black, but only a minimum habit strength for the response of discriminating.

Since it has been reported in the literature that a high degree of learning of one response reduces the likelihood of that response's interfering with the learning of retention of another response, it might be argued that the greater efficiency of Group E150 during reversal was due to a decrease in interference resulting from the overlearning of the first response. The data were thus analyzed to determine the amount of interference being transferred from the first to the second learning situations by the various groups. Interference was measured in terms of the daily blocks of five trials. A rat was considered to be responding in terms of the incorrect stimulus card, an overt erroneous intrusion, whenever it ran four out of five trials to this card regardless of its spatial position.

It will be noted from Table 2 that, as would be predicted by both the continuity and noncontinuity theories, those *S*s that had received the greatest number of reinforcements to the black continue to respond to that card for the greatest number of times after the reward relationship had been reversed. An analysis of variance carried out as above yields an F of 11.49, which is significant for 2 and 28 df at the 1% level of confidence or beyond. A comparison of the groups with each other provides a significant difference at least at the 5% level of confidence in every instance and the difference between Groups E150 and CO is significant at the .1% level ($t = 4.91$).

Discussion

Even though the group that was taken 150 trials past the criterion in original learning continued to run longer to the black stimulus card upon reversal, they also learned the reversal to white at a far more rapid rate than did either of the other groups. What happened was that very soon after the breakdown of the consistent response to the black card this group began to run consistently to the white card. Groups CO and E50, however, began to respond in terms of position habits and to stimuli other than the cards.

There are at least two possible interpretations that can be given these results. As was suggested earlier in the paper they may be accounted for in terms of *S*s learning to respond to a set of stimuli within the total stimulus complex—in this case the black and white stimulus cards. It may be well to consider how this response might be built up.

In the Y-discrimination apparatus *S* runs down the alleyway to the choice point and is likely, at least 50% of the time, to look at both the incorrect and the correct stimulus before being able to make a correct response. As the learning progresses, there develops a tendency on the part of *S* to stop at the choice point and shift his head in a manner that permits the stimuli from both cards to fall upon his retina. It is a response in some ways similar to that of "vicarious trial and error" reported by Muenzinger, Tolman, and their co-workers (**5, 7**). It is different in that VTE designates "the hesitating, looking-back-and-forth sort of behavior," while the response being described is a clear-cut "looking at" one stimulus card, "looking at" the other stimulus card, and immediately making a response to the correct card. If *S* makes such a response, it is reasonable to assume that upon being consistently followed by a reinforcing stimulus, food in this instance, it gains

in habit strength. The response, however, does not begin to occur consistently until the habits of avoiding the negative card and responding to the positive card are fairly well developed. Furthermore, when it does begin to occur consistently, increments in its habit strength are likely to be smaller than is the case of the response to the specific stimulus cards since the reinforcing stimulus does not immediately follow its being made. For these two reasons any substantial habit strength for the response of discriminating is built up only after a fairly large amount of overtraining in the original discrimination situation.

Two possible reasons may also be suggested as to why the extinction of this response is slower than the extinction of the response to the original positive stimulus card. The first is the fact that there is a greater opportunity for reactive inhibition to build up in the case of the latter response, and the second is the delay in nonreinforcement of the response of discrimination during the early reversal trials. It will be necessary to do further research to determine specifically why it is maintained after the response to the specific stimulus card has been extinguished, but its maintenance facilitates the learning of the reversal.[4]

The second interpretation of the results has been suggested by Spence[5] and follows directly from his theory of discrimination learning (6). It is possible that the rats upon reaching the criterion in the original discrimination problem still have differential position habit strengths. For those *S*s that are reversed immediately upon reaching the criterion this permits habit interference on the part of the stronger position habit in the learning of the new discrimination. The *S*s given an excessive amount of overlearning in the original problem, on the other hand, tend to have their position habits equalized. This occurs because of the smaller increments in habit strength on the part of the stronger habit with reinforcement, and because each position is reinforced an equal number of times during the overlearning period. Upon reversal these *S*s continue to run to the originally positive card which leads to a consistent extinction of that response and equal amounts of extinction with respect to the two position habits. Once the habit to the originally positive card is extinguished, *S* is left with equal position habits and the reinforcing schedule is such that the to-be-learned positive card is reinforced 100% of the time while each position is reinforced only 50% of the time. With no habit interference resulting from unequal position habits there is the opportunity for the reversal to be learned rapidly.[6]

It would seem likely that a differential test of the two interpretations is possible. In either event the behavior during the reversal, though it appears to be noncontinuous, can be accounted for only in terms of the continuity theory.

Summary

It was hypothesized that as a rat learns to make a specific choice dis-

[4] In the Harlow (1, 2) situation this response, if developed, is built up over a long series of different problems. The rapid shifting of problems may account for the slowness with which it is developed. Once having been built up, however, its extinction is likely to be even slower than is the case in the present study because of the differential amounts of negative transfer to the new reward relations in the two situations.

[5] Spence, K. W. Personal communication.

[6] It is possible that Spence's concept of "receptor-orienting acts" could be extended to account for the results of this experiment. There is undoubtedly some similarity between that concept and the one being offered in the present paper. In communications with Spence, however, he has suggested only the interpretation given above.

crimination, he is also learning a response of discriminating, i.e., learning to respond to a set of stimuli of which the specific stimulus is a member. A possible test of the hypothesis would be the demonstration of a more rapid learning of a discrimination reversal by *S*s that had received an excessive amount of overlearning in the original discrimination problem.

Three groups of 15 rats were trained to make a simple black-white discrimination in a Y-alley discrimination apparatus. All groups learned to 9 out of 10 correct responses with the black card positive, then Group CO was immediately reversed, Group E50 was given 50 overlearning trials before being reversed, and Group E150 was given 150 overlearning trials before reversal. Trials to reach a criterion of 9 out of 10 correct with white positive were determined for all groups.

Even though *S*s that were given 150 trials past the criterion in the original learning continued to run longer to the black stimulus card upon reversal, they also learned the reversal to white at a far more rapid rate than did *S*s of the other groups. Possible interpretations of these results are discussed.

References

1. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
2. Harlow, H. F. Analysis of discrimination learning by monkeys. *J. exp. Psychol.*, 1950, **40**, 26–39.
3. Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
4. Lawrence, D. H. Acquired distinctiveness of cues: II. Association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.
5. Muenzinger, K. F. Vicarious trial and error at a point of choice: I. A general survey of its relation to learning efficiency. *J. genet. Psychol.*, 1938, **53**, 75–86.
6. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
7. Tolman, E. C. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev.*, 1939, **46**, 318–336.

(Received December 6, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

MAGNITUDE OF REWARD AND ACQUISITION OF A BLACK–WHITE DISCRIMINATION HABIT[1]

JOEL E. GREENE[2]
University of Colorado

In his *Principles of Behavior* Hull (3) postulated that the amount of reward had a direct influence upon the asymptote of habit strength. Subsequent research (1, 12) yielded results which did not appear to be in strict conformity with predictions based upon this formulation. Hull (4) then modified his postulate set in order to take better account of these data. The revision states that magnitude of reward is the determinant of motive-incentive conditions (K) which are a constituent of reaction potential (sE_R), and which have a logical status within the system comparable to drive (D). The present investigation is an attempt to clarify the role of the reward variable in learning situations.

The investigations which apparently contributed most to the revised assumptions concerning the magnitude of reward were those of Crespi (1) and Zeaman (12). These two studies have in common the utilization of a running response in a straight alley. Crespi measured running speed, whereas Zeaman measured latencies. The findings of these two studies are similar. In general, the data show differences in the asymptotes of performance as a function of magnitude of reward. This is consistent with Hull's early formulation. However, under conditions of switched rewards two effects were noted: (*a*) Performance is modified in the direction of the switch almost immediately, and (*b*) the switched group does not approach the level of performance of a nonswitch group as an asymptote, but rather passes beyond it. Neither of these effects appears to be predictable from the postulate set propounded in *Principles of Behavior*, and it is these two findings which apparently induced Hull to modify his formulations.

Reynolds (5, 6, 7) has reported three studies in which the effect of different amounts of reward upon learning was investigated by training different groups under different size of reward conditions. The tasks to be learned included a lever-pressing habit, a spatial habit, and a black-white discrimination. In general, the findings indicated some difference between large and small reward groups in terms of response times. Measures of trials to criterion, number of trials to extinction, or percentage of correct trials yielded a statistically significant difference only in the spatial-discrimination problem.

The present investigation has been designed to show the influence of magnitude of reward upon excitatory tendencies toward particular cues. Two sizes of reward were presented in black and white goal boxes to two groups of white rats in preliminary training and the effects of this reward experience were measured in the acquisition of a black-white discrimination.

METHOD

Subjects.—The *S*s for this experiment were 48 naive albino rats from the colony maintained by the Department of Psychology at the University of Colorado. The *S*s, half males and half females, ranged in age from 60 to 90 days.

Apparatus.—Two pieces of apparatus were used in this experiment. The preliminary training apparatus consisted of a gray 8-in. alley, or starting unit, and two goal boxes, one white and the other black. A guillotine-type door sep-

[1] The present study is a portion of a thesis presented to the Faculty of the Graduate School of the University of Colorado in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer is indebted to Dr. M. P. Smith, who directed the investigation.

[2] Now at the University of Denver.

arated the alley from the goal box and there was also a rod placed across the top of the goal box so that a curtain could be inserted behind the door.

During the preliminary training, this apparatus was placed in the experimental room in such a manner that the starting unit occupied the same relative position as the choice point of the testing apparatus. A dim, even light was provided by a 24-in., 20-w. fluorescent tube suspended upside down 30½ in. directly over the apparatus. A shaded light was placed over the retaining cages where *S*s were kept between trials. These precautions were taken to minimize color preferences of the *S*s. The lighting conditions were identical for both the preliminary training and the testing trials.

The testing apparatus was essentially a duplication of the discrimination apparatus used by Grice (2). It consisted of a starting box leading into a narrow alley which terminated at a Y-shaped choice point. Two parallel straight alleys extending from the choice point were painted black and white, respectively. A third alley parallel to the other two permitted changing the right-left orientation of black and white by moving the alley unit. Guillotine-type doors were placed between the starting box and the narrow alley, and at the choice alleys, as well as between the alleys and the goal boxes. Curtains were inserted in front of the doors at the latter two locations. In every case the color of the curtains corresponded to the color of the alley or goal box.

Preliminary training.—After each *S* had been handled approximately 5 min. daily for one week, the preliminary training was begun. The handling was carried on at the same time as the experimental session. An important phase of this procedure was the feeding by hand of five Friskies pellets. This served the dual function of accustoming *S*s to the same type of food that was to be used as a reward as well as facilitating eating on the first trial of preliminary training.

During this period of handling, the feeding schedule was also established. The feeding regimen consisted of placing *S*s in individual feeding cages for 1 hr. at the same time daily resulting in a 22-hr. deprivation at the time of running. The feeding period followed the experimental session by 30 min. Water was available in the home cages at all times.

The preliminary training was conducted in the first apparatus described above. The *S* was placed in the starting unit and trained to run into the goal box for food. Each *S* was detained in the starting alley for at least 10 sec. and in the goal box for at least 30 sec. In no case was *S* removed from the goal box until the entire amount of food reward was consumed.

In order to add to the differential cues, which would acquire secondary reinforcing properties in the goal boxes, distinctive food cups were utilized. A circular white food cup was always in the white goal box, and a square black cup was always in the black goal box.

The *S*s were given five reinforced trials a day for six days with at least 15 min. between trials. At the end of the first ten trials, curtains were introduced into the apparatus.

Groups and conditions.—The *S*s were assigned at random to two groups in such a way that an equal number of males and females were in each training group. One group had small pellet in white box and large pellet in black box; the other had large pellet in white box and small pellet in black box. Thus, each group had an equal amount of experience with both pellet sizes, the difference in conditions being simply the color of goal box with which a given pellet was associated. This design was considered more appropriate because this procedure would tend to minimize the effect of a change in amount of food reward during the test trials. On any given day of training, only one size of pellet was used. For example, on the first day a given *S* would receive only the large-sized reward and on the next day he might receive only the small-sized reward. In order to control the possible effects of initial experience with rewards, half of each group was trained with the large pellet the first day, and half with the small pellet. Denoting the first day's pellet with the letter "A," the order of presenting pellet sizes by days for all groups was ABBABA.

The pellets were handmade from finely ground Friskies meal moistened with water. A constant check on the size of the pellets was provided by drawing random samples of 21 from each batch of pellets made after they had been dried for 48 hr. The means of these samples ranged from 233 mg. to 271 mg. for the large-sized pellets, and from 66 mg. to 85 mg. for the small. Pellets which were obviously too large or too small were discarded. All pellets were dried for at least 48 hr. before use.

Testing procedure.—On the day immediately following the completion of the preliminary training, test trials were begun. Each preliminary training group was divided at random so that an equal number of *S*s was tested under each of the reward conditions. Thus, half of each group received the large-sized pellet throughout the test trials and half received the small pellet.

The test procedure consisted of training *S*s to approach the black alley for food. No orientation to the new apparatus was necessary because of the relatively intensive preliminary training and the similarity of the two situations. The *S*s were given four free trials a day using the non-

TABLE 1

MEANS AND *SD*'S OF ERRORS DURING THE FIRST THREE DAYS OF TEST TRAINING AND OF TRIALS TO CRITERION FOR EACH REWARD CONDITION

Size of Reward and Location during Training		Size of Reward during Test	*N*	Errors during First 3 Days of Test		Trials to Criterion	
White Goal Box	Black Goal Box	Black Goal Box		Mean	*SD*	Mean	*SD*
Large	Small	Small	12	6.08	1.68	32.50	5.79
Large	Small	Large	12	5.67	1.50	29.33	4.79
Small	Large	Small	12	4.75	2.96	25.00	12.48
Small	Large	Large	12	4.25	1.71	25.08	10.29

correction method, with at least 15 min. intervening between trials. As was done during the preliminary training, *Ss* were detained in the starting box for 10 sec. and in the goal box for at least 30 sec. regardless of choice. The same distinctive food cups used in the preliminary training were present in the goal boxes of the discrimination apparatus. The white circular food cup, however, contained no food since it was on the incorrect side for all *Ss*.

Three precautions were taken to minimize the effects of position habits: (*a*) The correct alley was presented on both the right and left sides in an ABBA order each day; (*b*) half of the *Ss* had the correct alley on the right side for the first trial, and half had the correct alley on the left; and (*c*) the side of the correct alley on the first trial was alternated on successive days. All *Ss* were run to a criterion of nine out of ten correct trials.

RESULTS

Two measures of performance were taken during the test trials: (*a*) the number of white entries (errors) made during the first three days, and (*b*) the number of trials necessary to reach a criterion of nine out of ten correct choices. These two measures tend to supplement each other inasmuch as the error score reflects the influence of the preliminary training upon the initial portion of the test trials, whereas the trials to criterion data would seem to give some indication of the persistence of this influence throughout the acquisition of the black-white discrimination.

The mean and *SD* of each group for the number of errors made the first three days are given in Table 1. The data for males and females have been pooled in this table since the ensuing analysis of variance showed that sex was not a significant source of variance. In order to evaluate the possible interaction effects of preliminary training conditions and test conditions, the analysis-of-variance technique was utilized. The analysis performed was that recommended for a $2 \times 2 \times 2$ factorial design because the possibility of sex differences added another source

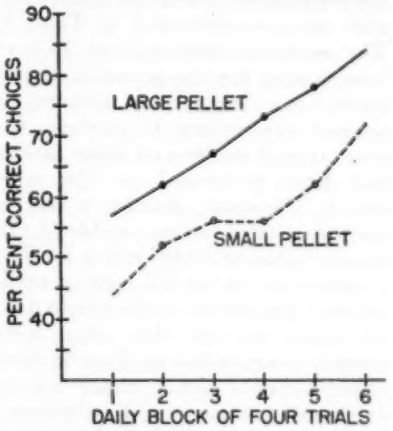


FIG. 1. B-W discrimination performance curves in terms of percentage of correct choices for groups pretrained with the large pellet in the black box and groups pretrained with the small pellet in the black goal box

of variance. Bartlett's test of homogeneity of variance yielded a nonsignificant chi square of 8.54 (7 *df*, $.30 > p > .20$).

The analysis of variance revealed that the only significant source of variability was the preliminary training. The *F* ratio, using the within-groups variance as error variance, was 5.07 ($p < .05$, 1 and 40 *df*). All other sources when compared with the within-groups variance yielded *F*'s less than unity. Fisher's *t* was computed for the difference between the means of all *S*s receiving the large-sized pellet in the black goal box during preliminary training, and of all *S*s receiving the small-sized pellet in the black box. Using the within-groups variance as the best estimate of the population variance, *t* was 3.21 ($p = < .01$, $df = 40$).

Figure 1 shows the percentage of correct choices per day for the above two groups for the first six days of test training.

The mean and *SD* for each group for the number of trials to reach criterion are also presented in Table 1. The results for males and females have been pooled for the purposes of this table. For comparable reasons, it seemed appropriate to perform the same type of analysis on these data as had been performed on the error scores. However, Bartlett's test of homogeneity of variance yielded a chi-square value of 18.40, which for 7 *df* is significant at the 2% level of confidence. Inspection of the distribution of scores showed that they were roughly symmetrical so that the heterogeneity of variance could not be attributed to skewness. Furthermore, those groups that had the large-sized pellet in the black goal box during preliminary training were consistently more variable than those groups that had the small pellet in the black goal box. An indication of this tendency is provided by the *F* of 4.22 computed between the total variance of the large-pellet groups and the total variance of the small-pellet groups. With each variance based on 23 *df*, a value of only 2.70 was needed for significance at the 1% level of confidence.

A critical ratio was computed for the difference between the means of the two groups representing the two different-sized pellets during the preliminary training because this statistic does not assume homogeneity of variance. The ratio yielded a value of 2.3; $p = .02$.

Because of the possible effects of the difference in variability between the two groups in question, another method was used to verify the tendency of the group receiving the large pellet in the black goal box during the preliminary training to reach criterion faster than the group getting the smaller reward. A chi square was computed for the frequencies of each pellet-sized group reaching criterion above and below the median number of trials required by all *S*s. Chi square computed in this manner was 5.32, which for 1 *df* has a *p* value of between .05 and .02.

It appeared to be useful also to evaluate the possible effects caused by the size of pellet used during the test conditions with the preliminary training disregarded. Consequently, the same analyses were applied to these data as were used in the appraisal of the effects of preliminary training. The *F* ratio of the variance of all *S*s receiving the small pellet to the variance of all *S*s getting the large pellet was 1.58. A value of 2.00 was needed for *p* to equal .05. The critical ratio of the differences between the means of these two groups was .57, which falls considerably short of significance.

A chi square, based on the frequency

with which *Ss* in the two groups differentiated by the pellet size during the test conditions fell above or below the median performance of all *Ss*, yielded a value of 1.33, which for 1 *df* has a *p* value falling between .30 and .20.

Discussion

The results of this experiment clearly demonstrate that those *Ss* that had experienced a large food pellet in the black goal box during preliminary training tended to make fewer choices to the white side and take less trials to reach criterion to the black side in a black-white discrimination problem than those *Ss* that had experienced a large food pellet in the white goal box during preliminary training. Furthermore, the size of the food pellet used as a reward in the discrimination situation had no significant effect upon performance.

In interpreting these findings, it is reasonable to assume that differential excitatory tendencies to black and white cues were established during the preliminary training and that these tendencies determined, in part, choices in the test situation. This is demonstrated by Fig. 1 which clearly shows that those *Ss* having had the large pellet in the black goal box during preliminary training started out at a higher level of performance in the test situation than those having had the large pellet in the white goal box. This superiority was maintained throughout the first six days of test training.

The effect of the preliminary training is also illustrated in the fact that the group which had the large pellet on the black side during the preliminary training and the small pellet during the B–W discrimination made fewer errors and reached criterion faster than the group which had had the reverse reward conditions in the preliminary training and the large pellet during the test. This appears to be somewhat in contrast to those studies which showed a more or less immediate change in performance as a function of change of reward. The procedure of giving all *Ss* equal experience with both-sized rewards may have served to minimize the effects of reward-pellet magnitude in the test situation. However, a pilot study conducted by the writer in which two groups were pretrained to the white goal box only, one with 75 mg. reward and the other 230 mg. reward showed essentially the same findings, i.e., the B–W discrimination problem was more sensitive to the pretraining reward conditions than to the size of reward during the test. This despite the fact that in this pilot study half of the *Ss* experienced a change to a reward with which they had had no prior experience. Perhaps the major factor which served to minimize the effectiveness of test-pellet size is the tendency of error scores or trials to criterion measures to be less sensitive to changes in reaction potential than measures of latency or response times.

In any case, the results indicate that differential excitatory tendencies developed in the separate presentation of cues as a function of magnitude of reward are capable of persisting in a situation where the cues are presented simultaneously.

Data reported by Reynolds indicate that the magnitude of reward is an effective variable in the acquisition of a simple spatial habit (**6**), and is not effective in the acquisition of a black-white discrimination habit (**5**). The findings in the latter study are in direct contrast to those reported here. A possible explanation for this discrepancy is suggested by Spence's analysis of discrimination learning (**10**).

Reynolds utilized no preliminary training procedure comparable to that employed by the writer. This resulted in the simultaneous development of inhibitory effects and approach tendencies. Furthermore, the inhibitory effects would tend to be equal for both reward groups. Since the making of a particular discrimination response is a joint function of both weakening and strengthening of approach tendencies, it seems tenable to assume that the equalization of inhibitory effects for both groups tended to obscure any difference which might have existed attributable to variation in the reward variable. The fact that the correction method of training was used probably accentuated the role of these inhibitory effects. In the present experiment, the preliminary training emphasized the development of approach tendencies and minimized the possible contaminating effects of nonreinforcement.

The results of this study are consistent with the theoretical formulations proposed by Hull (**3**), Spence (**11**), and Seward (**8, 9**). The inability of the experiment to test these explanations differentially arises from the fact that in varying pellet size, not only was the nutritive value of the reward varied, but also the magnitude of the eating response was different for the two pellet size conditions. Consequently, it cannot be certain whether the data obtained can be attributed to the differential development of ${}_{S}H_{R}$ through different amounts of reinforcement (**3**) or to a difference in excitatory tendencies resulting from a difference in strength of consummatory response (**8, 9, 11**). It would appear, however, that Hull's 1949 revision without Spence's amplification would have some difficulty in treating these data, since K is assumed to be a function of magnitude of reward only, and is hypothesized to shift promptly in value under conditions of changed reward. This formulation would predict that the reward-pellet size employed during the testing situation would determine the value of K and hence the excitatory tendency, irrespective of the reward conditions during preliminary training. Spence's interpretation overcomes this difficulty by assuming that the mechanism underlying K is the fractional anticipatory goal response and its interoceptive cue. Thus, assuming that the large reward was presented in the black goal box and the small reward in the white goal box, S_B became conditioned to a bigger r_G than did S_W. Therefore K is greater in the case of S_B and, consequently, the approach tendency to S_B is stronger.

Summary and Conclusions

Two groups of rats were given a series of 30 preliminary training trials in a short runway terminating in a colored goal box. One group received 15 reinforced trials with a small food pellet in a black goal box and 15 reinforced trials with a large food pellet in a white goal box. The other group was given 15 reinforced trials with a large pellet in a black goal box and 15 reinforced trials with a small pellet in a white goal box. Immediately following the completion of the preliminary training, all *S*s were trained in a black-white discrimination situation with black positive. Half of each of the preliminary training groups were rewarded with the large pellet and half with the small pellet in this testing situation.

The results of this experiment were as follows:

1. The group receiving the large pellet in the black goal box and the small pellet in the white goal box

made significantly fewer errors during the first three days of test training and learned to criterion more rapidly than the group pretrained with the small pellet in the black goal box and the large pellet in the white goal box.

2. Neither the error measure nor the trials-to-criterion measure resulted in a significant difference between the group rewarded with the large pellet during test training and the group rewarded with the small pellet during test training.

It was concluded that differential excitatory tendencies, developed to separately presented cues as a function of magnitude of reward, are capable of persisting in a situation where the cues are presented simultaneously.

References

1. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
2. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
5. Reynolds, B. The acquisition of a black-white discrimination habit under two levels of reinforcement. *J. exp. Psychol.*, 1949, **39**, 760–769.
6. Reynolds, B. Acquisition of a simple spatial discrimination as a function of the amount of reinforcement. *J. exp. Psychol.*, 1950, **40**, 152–160.
7. Reynolds, B. Resistance to extinction as a function of the amount of reinforcement present during acquisition. *J. exp. Psychol.*, 1950, **40**, 46–52.
8. Seward, J. P. Experimental evidence for the motivating function of reward. *Psychol. Bull.*, 1951, **48**, 130–149.
9. Seward, J. P. Delayed reward learning. *Psychol. Rev.*, 1952, **59**, 200–201.
10. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
11. Spence, K. W. Theoretical interpretations of learning. In C. P. Stone (Ed.), *Comparative psychology.* New York: Prentice-Hall, 1951. Pp. 239–291.
12. Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received December 8, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

ANXIETY AND TASK AS DETERMINERS OF VERBAL PERFORMANCE [1]

CHARLES K. RAMOND
State University of Iowa

Certain recent investigations (**4, 5, 6, 7, 8**), concerned with the effect of motivational level on human performance in various learning situations, have attempted to manipulate drive by the selection of subjects, rather than by the usual method of varying certain antecedent conditions (e.g., degree of noxious stimulation, or degree of stress in a situation). Specifically, this method has consisted in selecting two groups of *Ss* on the basis of their having obtained extreme scores on a test of manifest anxiety developed by Taylor (**6**). The assumption was made that an *S*'s effective drive strength is a direct function of his score on this test of manifest anxiety. Thus *Ss* with high scores on the Taylor anxiety scale were assumed to have high drive strength, while those with low scores were assumed to have low drive strength.

If one assumes, as does Hull (**2**), that response strength is a multiplicative function of drive strength and habit strength, i.e., that $R = f\ (D \times H)$, then it follows that the higher the level of D, the greater will be the strength of any given response tendency. Confirmation of this set of assumptions was obtained in the initial study by Taylor (**6**), who found that anxious (high-drive) *Ss* exhibited a markedly higher level of performance in an eyelid-conditioning situation than did nonanxious (low-drive) *Ss*. A subsequent eyelid-conditioning study by Spence and Taylor (**5**) confirmed this finding.

A higher motivational level, however, will not always result in superior performance. As Taylor and Spence (**7**) recently pointed out, the effect of added drive upon performance will depend upon the nature of the task to be performed. If the task is a simple one in which there is but a single S–R tendency, then higher drive will lead to a higher level of response. But in more complex learning situations, which involve a hierarchy of competing responses, the effect of a higher drive level will depend upon both the number of responses in this hierarchy, and the relative habit strength of the correct response as compared with the habit strengths of the incorrect responses. In a situation in which there are many competing S–R tendencies, particularly in a situation in which the correct response is initially weaker in habit strength than one or more incorrect responses, it can be shown that the effect of higher drive strength will be opposite to that found in the conditioning experiments, i.e., it will lead to *poorer* performance. This prediction follows from the assumption that drive multiplies indiscriminately the habit strengths of all responses, thus, in this competing-response situation, increasing the amounts by which the excitatory strengths of the incorrect responses exceed that of the correct response.

This theoretical expectation has also been confirmed in more complex learning situations. Montague (**4**) found that anxious *Ss* were significantly inferior to nonanxious *Ss* in the

[1] The writer is indebted to Prof. Kenneth W. Spence, who directed the investigation. This study was done under Contract N9onr-93802, Project NR 154-107, between the State University of Iowa and the Office of Naval Research.

rote learning of a list of nonsense syllables chosen in such a way as to produce strong competing responses. But the performance levels of the groups were reversed on a list in which an attempt was made to minimize the strength and number of competing response tendencies. Taylor and Spence (7) likewise found that anxious *Ss* were inferior to nonanxious *Ss* in a verbal maze situation.

The present study investigated the relationship between drive level, defined in terms of scores on the Taylor manifest-anxiety scale, and performance in a simple trial-and-error learning situation in which an attempt was made to control the number and strength of the competing responses. This situation consisted of a series of stimulus words to each of which *S* could respond with either of two given response words. These response words were selected so that one was relatively high in the initial response hierarchy and the other relatively low. Thus *Ss* chose, at each presentation of a stimulus word, between a response connected to the stimulus with relatively high habit strength, and a response connected to the stimulus with little or no initial habit strength. Performance was observed on two kinds of presentations: In the first, the stronger of the two responses was correct (reinforced) while the weaker was incorrect (not reinforced). In the second, the weaker of the two responses was correct and the stronger incorrect.

Method

Subjects.—The *Ss* used in this experiment were selected from an undergraduate psychology course on the basis of their having obtained extreme scores on a modification of the original Taylor manifest-anxiety scale (6). This test consisted of items taken from the Minnesota Multiphasic Personality Inventory, items which had been judged by five clinical psychologists to be indicative of a carefully defined behavior syndrome, manifest anxiety. The modified test uses those 50 of the original 65 items which correlated highest with the total anxiety score on the original scale. The Lie Scale of the MMPI was used to detect false extreme scores, an L-score in the upper decile eliminating an *S* from consideration for the experiment.

All *Ss* were naive with respect to the experimental task and were unaware of the reason for their selection, having been told that they were chosen by random sampling procedures.

The anxious group consisted of 33 individuals whose scores on the anxiety scale fell in the upper 15% of a distribution of approximately 2000 scores previously made by undergraduates who had taken the test. The nonanxious group was composed of 40 individuals whose scores fell in the lower 20% of that distribution.

Apparatus.—A Hull-type memory drum was mounted across two tables and set in an 18 × 12-in. opening of a flat-black, fiberboard screen. The exposure aperture of the drum was 3 × 1 in. A large twin-bulb fluorescent lamp placed against the top of the opening provided illumination. The position of this lamp was adjustable to the sitting height of the *S* so that sight of the aperture was never obstructed.

A stimulus presentation consisted of the simultaneous appearance in the aperture of three two-syllable adjectives, one on the left and two on the right, as follows:

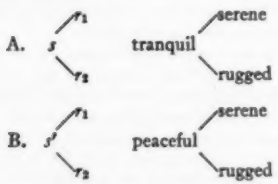
Stimulus Word	*Response Words*
tranquil	serene
	rugged

Four different orders of a list of 16 such presentations were typed in lower-case letters on an endless cloth tape. Parallel to this test list was typed in similar fashion a practice list of two-syllable nouns. A sliding mask on the aperture allowed *Ss* to see only one list at a time.

Each presentation was exposed in the aperture for 2.4 sec., with a .25-sec. interval between exposures. The interlist interval was 5.3 sec. These odd intervals resulted from the following modifications of the Hull-type drum: (*a*) removal of the shutter mechanism, and (*b*) division of the standard 8-sec. revolution period of the drum into three equal exposure intervals instead of four. The former change was effected because the learning method involved choice rather than anticipation of the correct response; the latter because a preliminary study indicated that the 2-sec. exposure interval was too short to permit learning within a limited number of trials.

It is evident in the above sample presentation that one of the two response words is associated much more closely with the stimulus word than is the other response word. This was the case on each presentation, the assumption being that the habit strength of one S–R connection in each case greatly exceeded that of the other. Habit

strength here was thought of as being equivalent to the *association value* holding between the words, as defined by Haagen (1). The following paradigm will help illustrate how the presentations were selected:



A and B represent separate presentations. In A, r_1 is correct, while in B, r_2 is correct. Eight such pairs of presentations comprised the test list. Thus *S* saw the same pair of response words *twice* in any given order of the list, but the response word which was correct the first time he saw the pair was *not* correct the second time. Hence *S* was forced to discriminate the response pairs according to their different stimulus words. For example, *S* had to learn that whenever he saw "tranquil" with "serene" and "rugged" as possible responses, *serene* was the correct response to make; but when he saw "peaceful" with "serene" and "rugged," *rugged* was the correct response.

The appearance of the same pair of response words twice within a given order of the list (with one response being correct on one presentation and the other response correct on the other presentation) prevented *S* from: (*a*) learning the correct responses serially; (*b*) learning the correct responses without referring to the stimulus words.

Each pair of presentations was selected from Haagen's data according to the following criteria: (*a*) r_1, of five words chosen for their similarity to *s*, was scaled "closest in associative connection" to *s* by 80 judges; (*b*) *s'*, of five words chosen for their synonymity to *s*, was scaled "most similar in meaning" to *s* by 80 judges; (*c*) r_2 was chosen so that *no* associative connection between it and either *s* or *s'* was apparent.

Procedure.—After *S* was seated before the apparatus and the lamp adjusted, instructions were read aloud by *E*. *S* was told he would see a succession of presentations of "one stimulus word on the left and two response words on the right," and that his task was to learn, in each case, which of the two words on the right was the correct response to make. This he was told to do by saying aloud his choice at each presentation, thereupon discovering whether it was right or wrong by *E*'s reply: "right" when correct, and no reply when wrong. The *E* was seated behind the screen and recorded *S*'s responses.

The learning task proceeded as follows: (*a*) 16 warm-up trials on the practice list (two-syllable nouns, arbitrarily selected); (*b*) a 2-min. rest; (*c*) 16 trials on the test list; (*d*) a 2-min. rest, during which time *E* engaged *S* in conversation to deter rehearsal; (*e*) 16 more trials on the test list, making 32 test trials in all.

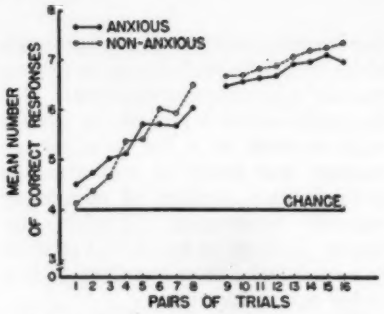


FIG. 1. Performance curves for those presentations in which the stronger response was correct

RESULTS

Performance curves of anxious and nonanxious *S*s are compared in Fig. 1 and 2, the former representing those presentations in which the stronger response was reinforced and the latter representing those presentations in which the weaker response was reinforced. Each point on both graphs stands for a group mean of the number of correct responses made on a pair of successive trials. Since each trial consisted of eight presentations in which the stronger response was correct and

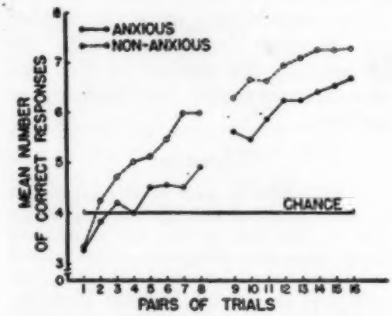


FIG. 2. Performance curves for those presentations in which the weaker response was correct

TABLE 1

MEAN NUMBER OF STRONG AND WEAK CORRECT RESPONSES MADE BY ANXIOUS AND NONANXIOUS SUBJECTS

Correct Response	Anxious (N = 33)		Nonanxious (N = 40)	
	Mean	*SD*	Mean	*SD*
Strong	193.63	23.04	196.83	43.34
Weak	165.24	36.00	190.02	39.98

eight in which the weaker response was correct, the maximum number of correct responses possible for either type of presentation was eight, with chance level of response, denoted by the horizontal line in the figures, being four.

Table 1 shows the mean number of correct responses for each group under both conditions. Since this variable was not distributed normally and also since the variance of this measure was heterogeneous between groups, the Mann-Whitney U test (3), which requires no assumptions about these parameters of the distributions, was employed. On those presentations in which the weaker response was correct, the nonanxious *Ss* responded correctly significantly more often than did the anxious *Ss*, the obtained value of the statistic permitting rejection of the null hypothesis of no difference between the distributions at well beyond the 1% level of confidence (p = .0016). On those presentations in which the stronger response was correct, however, no significant difference between the performances of the two groups was found (p = .17).

DISCUSSION

Prerequisite to the evaluation of the theoretical implications of the foregoing results is an unambiguous statement of deducible expectation with respect to both experimental situations: that in which the stronger response was correct, and that in which the weaker response was correct. In the latter case, the performance of the nonanxious *Ss* should have been superior to that of the anxious *Ss*, and such a result was obtained. The prediction of this result is based upon the assumption that effective drive multiplied the habit strengths of both responses. Since in this situation (weaker correct) the incorrect response is higher in the initial hierarchy, this would have increased the amount by which its excitatory strength exceeded that of the correct response. The effect of this was to lower the probability of occurrence of the correct responses, thus increasing the number of errors and retarding learning. Since this retardation is directly proportional to the amount of the abovementioned effective drive, the performance of the anxious *Ss* was inferior to that of the nonanxious *Ss*.

The theoretical prediction concerning the performance of the two groups in that situation in which the strong response was correct, however, is complicated in the present study by the action of stimulus generalization. It may be predicted that the anxious group would perform better (make a higher number of correct choices) at the very beginning of learning. But as soon as the *Ss* began to learn the weaker response in the situation in which the weaker response was correct, then this learning generalized to the synonymous stimulus word and the weaker response gained strength in the situation in which the stronger response was correct. Because the anxious group had a higher D level, this would mean that there was a greater likelihood for the excitatory strengths of these incorrect weak responses to be superthreshold for the anxious rather than for the nonanxious group. The operation of oscillation would thus have provided for the occasional dominance of these "weaker" S–R tendencies.

The interaction of all these opposing factors precludes the formulation of any precise deduction. All that may reasonably have been predicted was that the anxious group should have been superior to the nonanxious group early in training, and that this superiority should have decreased as training progressed. Figure 1 shows that such was the case, performance of the anxious *Ss* becoming, in fact, inferior to that of the nonanxious *Ss* after about four pairs of trials.

Summary and Conclusions

This experiment investigated the effect of differences in motivational level upon performance in a verbal learning situation as a function of the relative strengths of the correct and incorrect responses in the initial hierarchies.

Motivational level was assumed to be positively related to scores on the Taylor manifest-anxiety scale, and was varied by selecting two groups of *Ss*, one anxious (high-drive) and one nonanxious (low-drive), on the basis of their having obtained extreme scores on this test. The relative strength of the correct response in the initial response hierarchy was controlled in two ways: (*a*) the hierarchy was restricted by forcing *Ss* to choose, on each presentation, one of two responses; (*b*) one of these two available responses was made stronger than the other on the basis of scaling techniques.

Specifically, each presentation consisted of three two-syllable adjectives, one on the left and two on the right, which appeared simultaneously in the aperture of a Hull-type memory drum. Sixteen such presentations constituted a trial. In eight of the presentations, the stronger response was reinforced; in the other eight, which were randomly interspersed with the first eight, the weaker response was reinforced. Thirty-two learning trials were given, 16 before and 16 after a 2-min. rest interval.

The following results were obtained: (*a*) nonanxious *Ss* responded correctly significantly more often than did anxious *Ss* on those presentations in which the weaker response was correct; (*b*) there was no significant difference between the performance of the anxious and nonanxious *Ss* on those presentations in which the stronger response was correct.

These results were interpreted in relation to Hull's assumption that drive strength (*D*) multiplies all stimulus-response habits. It may be concluded that the effects of differences in drive strength depended, in the present experiment, upon whether the correct response was the stronger or the weaker in the hierarchy.

References

1. Haagen, C. H. Scaling the principal attributes of words to be used in experiments in verbal learning. Unpublished M.A. thesis, State Univer. of Iowa, 1942.
2. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
3. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
4. Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.
5. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.
6. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.
7. Taylor, J. A., & Spence, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.
8. Wenar, C. Reaction time as a function of manifest anxiety and stimulus intensity. Unpublished Ph.D. dissertation, State Univer. of Iowa, 1951.

(Received December 11, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

A PSYCHOPHYSICAL AND ELECTROPHYSIOLOGICAL STUDY OF LIGHT ADAPTATION

ROBERT M. BOYNTON [1]
University of Rochester
AND M. HOWARD TRIEDMAN [2]
Brown University

Light adaptation may be defined as a change in sensitivity of the dark-adapted eye as a result of stimulation by light. Sensitivity, in turn, is usually defined as the reciprocal of a threshold measure. There are, however, two fundamentally different ways in which threshold (and therefore sensitivity) has been measured in the evaluation of the process.[3]

Recovery method.—Light adaptation is measured by noting some aspect of the recovery in sensitivity of the eye following pre-exposure to an adapting stimulus. A typical set of results might show absolute thresholds, determined, for example, 5 sec. after the cessation of an adapting stimulus, plotted as a function of the duration of the adapting stimulus.

Intensity-discrimination method.—During the presentation of an adapting stimulus, and at specified times following its onset, intensity-discrimination thresholds (ΔI) are determined. A typical result simply shows ΔI as a function of time in the light.

Data obtained by the recovery method are best given a photochemical interpretation. An outstanding illustration is given by Wald and Clark (11) and is supported by the electrophysiological data of Hartline and McDonald (6).

Data obtained by the intensity-discrimination method do not easily allow a photochemical interpretation. One of the necessary consequences of a simple photochemical theory is that sensitivity should continuously decrease following the onset of an adapting stimulus, until a steady state is reached. However, sensitivity, as determined by the intensity-discrimination method, shows an increase (rather than a decrease) during part of the adaptation process. At photopic adapting levels, this increase has been shown to begin during the first fraction of a second of adaptation by Crawford (5), and has been found to continue for one to three minutes by Baker (1), after which time Baker's curves show a reversal and a slow decrease in sensitivity begins. Baker interpreted his results by suggesting that:

> In early adaptation, the near-maximum rate of response of the receptor [to the adapting stimulus] limits its ability to respond to an added intensity, resulting in a high increment threshold. As the receptor adapts, the limit due to this factor declines until a point is reached at which the photochemical process becomes limiting, when the increment threshold is at a minimum and subsequently rises again (2, p. 279).

Crawford's data suggest that the sensitivity changes noted by Baker, which were on the order of .3 log unit, are relatively small compared to a

[1] These experiments were conducted in the Psychological Laboratories of Brown University, under contract between the Office of Naval Research and Brown University. The authors wish to thank Dr. Lorrin Riggs for making his facilities available, and for his critical reading of the manuscript.

[2] Mr. Triedman is now a student at the College of Physicians and Surgeons, Columbia University.

[3] There is a third method, the binocular-comparison method, which will be discussed briefly later in the paper.

larger decrease in sensitivity which he found at the beginning of the adaptation process. A surprising feature of Crawford's results was that the threshold to a test flash was raised above the dark-adapted level when the flash was delivered as much as .1 sec. *before* the onset of an adapting stimulus. When ΔI was determined with simultaneous onset of the adapting stimulus and the test flash (now an intensity increment), sensitivity had already dropped some 3 to 4 log units (depending on adapting luminance). Crawford interpreted these rather surprising results as follows:

> There seem to be two possible explanations. Either the relatively strong conditioning stimulus overtakes the weaker test stimulus on its way from retina to brain and interferes with its transmission; or the process of perception of the test stimulus, including the receptive processes in the brain, takes an appreciable time of the order of 0.1 sec., so that the impression of a second (large) stimulus within this time interferes with perception of the first stimulus (5, p. 285).

In the present study, psychophysical and electrophysiological evidence obtained at scotopic levels is presented which appears to substantiate, under new conditions, both the results obtained and interpretations suggested by Baker and Crawford. The psychophysical response index is the intensity-discrimination threshold, and the electrophysiological index is the *b*-wave component of the human electroretinogram (ERG).

Apparatus

Stimulating system.—Figure 1 gives a schematic view of the stimulating apparatus, which is best conceived as two highly similar optical systems. One of these, shown on the right, provided the adapting stimulus and will be referred to as the *adapting system.* The other, on the left, provided a brief test flash, and will be designated the *test-flash system.* The arrangement was such that the filament images of two identical 50 cp lamps (SL_1 and SL_2) were superimposed on the pupil of *S*'s eye. This was a double

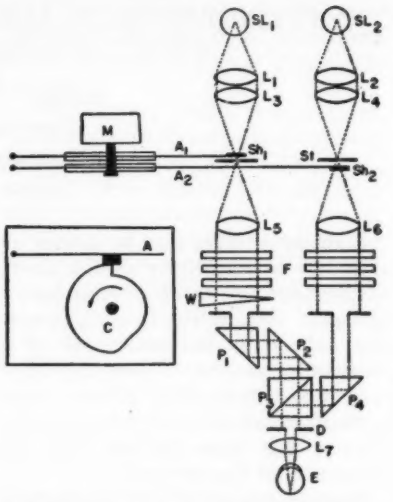


Fig. 1. Diagram of the stimulating system. The optical system on the left is the test-flash system, and that on the right, the adapting system. The component parts are described in the text.

"Maxwellian view" optical system, where the stimulus field, as seen by *S*, was limited in size by diaphragm *D* and appeared as if in the plane of the last lens of the optical system (L_7). The two optical systems stimulated identical retinal areas.

Light from the lamp was collimated (i.e., between L_1 and L_3) and then brought to a focus at stop *St*, where a first real image of the lamp filament appeared. At this point, shutter leaves, Sh_1 and Sh_2, interrupted the beams in order to control the temporal sequence of stimulus presentation. The shutter leaves were carried by arms A_1 and A_2. The position of the arms, in turn, was determined by the position of two cams (*C*), one of which is shown in side view in Fig. 1. The cams were driven by a common shaft attached to a 4 rpm synchronous motor. The timing of the stimuli was controlled by the relative position of these two cams. When A_2 fell, carrying Sh_2 with it, the adapting stimulus came on and remained on for a second or two until turned off by *E* with an auxiliary shutter (not shown). When A_1 fell, a circular aperture in Sh_1 swept across the circular stop aperture (of the same size, 2.5 mm. in diameter) and provided the test flash.

Lenses L_5 and L_6 again collimated the light, and neutral filters (*F*) were interposed here as

shown. Also indicated in Fig. 1 is the prism arrangement which was utilized to bring the two beams together.

The lamps were operated from storage batteries at a constantly monitored current of 10 amp. (about 6 v. across each bulb), slightly below the peak capacity of the bulbs. The color temperature of the bulbs was not measured, but was in the neighborhood of 2750° K.

A small red fixation point was provided so that *S* received the stimuli, in the psychophysical experiments, centered at 10° on the horizontal axis of the nasal retina of the right eye. The stimuli subtended a visual angle of 3°. In the ERG experiments, the stimulus was centered about the fovea and subtended 7°. Head position was controlled with a biting board.

Recording system.—The methods used for recording the ERG from human *S*s were the same as described elsewhere (**4**). The following are the important points: (*a*) The active electrode was mounted on the corneal bulge of a contact lens worn in *S*'s right eye, a method first devised by Riggs (**7**). (*b*) The indifferent electrode was in contact with *S*'s forehead. (*c*) The potential difference between these two electrodes (which have been shown to be effectively in contact with the front and back of the eyeball) was supplied to the input of a direct-coupled amplifier. (*d*) The amplifier output was delivered to a loop galvanometer containing a small mirror. Light was reflected from this mirror onto a continuously moving strip of photosensitive paper which provided a permanent record of the changes in electrical potential between the two electrodes on *S*. Optical arrangements were also made so that the duration of exposure provided by each optical system was indicated, by a trace, on the paper. The latter will be referred to as the *signal system.*

Calibrations.—The relative luminance supplied by each system was varied by means of Wratten "neutral" filters, located as described, and a calibrated optical wedge (*W*). In the psychophysical experiments, a "blue" interference filter was located immediately behind diaphragm *D*.[4] This filter had a dominant wavelength of 424 mμ, and a scotopic density of about 2.5 log units, at the color temperature of the lamp used. (This value was determined by direct experiment, by obtaining absolute thresholds with, and without, the blue filter in the system.) The photometry of nearly monochromatic flashes of light near absolute threshold is fraught with difficulty, and no exact photometric specification of the luminance level was attempted for this study, where relative values seemed of primary importance.

The "sensitivity" values given in *Results* represent the density of neutral filters required for a threshold measure to be obtained. Higher sensitivity is thus represented by greater filtering at threshold, and the plot is an upside-down mirror image of a log-luminance threshold representation. In the electrical recording experiments to be described, the blue filter was not used in conjunction with the neutral filters.

With both shutter leaves removed, the amount of light transmitted by the two systems was equal. The characteristics of the shutters were determined by noting the response of a photocell placed in the position normally occupied by *S*'s eye. This also served to calibrate the signal system. The rates of onset of the test and adapting stimuli appeared almost identical. The shape of the entire voltage vs. time curve, as observed on the CRO for the test flash, was like that of a positively skewed normal distribution curve. Flash onset occupied .013 sec., and offset, .020 sec. At peak, only about 80% transmission occurred. Figure 2 shows a set of such calibrations, and indicates the kind of signal traces from which time measurements were made.

The relative positions of the cams which provided the interval between adaptation onset and test-flash onset could not be preset with accuracy to much better than .1 sec. However, by recording the signal traces, it was possible after each experimental session to determine, within .01 sec., what each time interval had been. Thus, the recording procedure was always carried out, even in the psychophysical experiments, where there was otherwise no reason to run the paper. Experiment showed that once the two cams were adjusted for a given interval, the interval remained constant to within much less than .01 sec.

Procedure

Psychophysical experiment.—Data were secured from two thoroughly practiced *S*s (the authors). The data resulted from two 3-hr. sessions with each *S*, each session being conducted on a separate day. The procedure during each experimental session was as follows: (*a*) The *S* was dark adapted for a minimum of 30 min. (*b*) At the termination of dark adaptation, a measure of absolute threshold was secured. This was obtained with a descending method of limits, with individual flashes 15 sec. apart in each series. Light from the adapting system was blocked off. (*c*) For a given interval, from the

[4] This filter has been described in detail elsewhere (**8**), where it was given the arbitrary designation "M." The filter was used to insure that only rods would be involved in the investigation, so that the data could be compared to the scotopic *b*-wave data. The *S*s perceived no color at any time during the experiment.

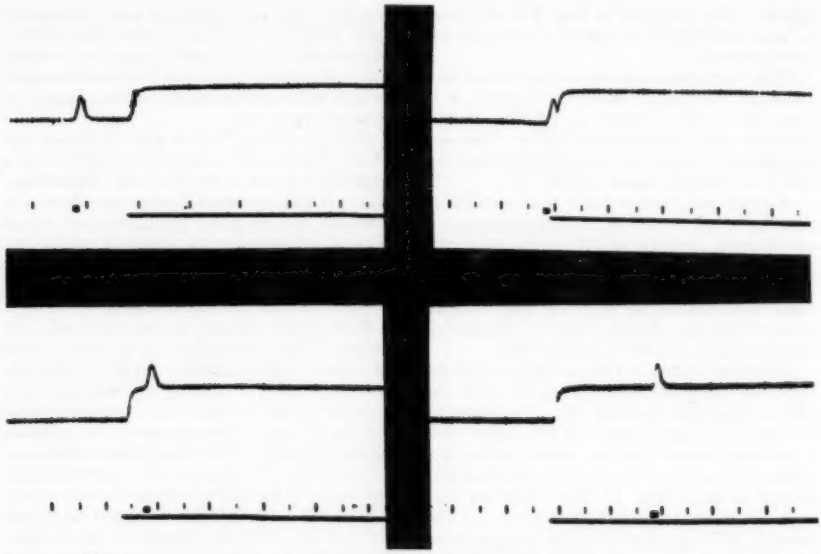


FIG. 2. Calibration records. The upper traces show the voltage output of a photocell, placed in the position normally occupied by *S*'s eye (see Fig. 1). The lower traces are those delivered to the recording paper by the "signal system," which also indicates the presence of the stimuli. The test flash is indicated by the short upper trace, and the adapting stimulus by the long lower trace. The heaviest vertical marks are .1 sec. apart.

onset of the adapting stimulus to the onset of the test flash (hereafter referred to simply as "adapting interval"), an intensity-discrimination threshold was obtained. The *S* was asked to judge the presence or absence of an intensity increment. "Blank" trials were occasionally presented, where the test-flash system was blocked off. The same descending method of limits described above was employed at all adapting intervals. (*d*) The procedure given in (*c*) was repeated for other adapting intervals. The sequence of adapting intervals was counterbalanced between the two experimental sessions, and included negative intervals (where the test flash preceded adaptation onset). (*e*) The absolute threshold was determined again as in (*b*). Three log units of neutral density filtering were used in the adaptation system at all times. Since absolute threshold was found with about four log units of filtering in an equivalent system, the luminance of the adapting stimulus was only about one log unit above absolute threshold (actually somewhat more, because of its extended duration).

Electrical recording experiment.—Data were obtained from two experienced and unusually "steady" *S*s. These were not the same *S*s used in the psychophysical experiment. Altogether, there were six 2- to 3-hr. sessions with each *S*. The procedure during each experimental session was as follows: (*a*) The *S* dark adapted for a minimum of 30 min. (*b*) At the termination of dark adaptation, the apparatus was adjusted for a particular adapting interval, and the following three types of records were obtained: response to adapting stimulus alone; response to test flash alone; response to adapting test-flash sequence. (*c*) Procedure (*b*) was repeated for other test-flash luminances, but at the same adapting luminance.

In subsequent sessions, this procedure was repeated for different adapting intervals. Two log units of neutral density filtering were used in the adapting system at all times. The blue filter used in the psychophysical experiment was not employed. The luminance of the adapting field was therefore about 3.5 log units higher than that used in the psychophysical experiment.[5]

[5] This may be confusing to those not familiar with earlier experiments (**3, 4**) where it was shown that the *b*-wave to small-area stimuli arises mainly from stray light stimulation of the entire retina. Had very large stimulus fields

Results

Psychophysical determinations.—The results obtained from one *S* are shown in Curve *B* of Fig. 3. Note the time base upon which these data are plotted: The entire sequence of sensitivity changes illustrated took place in about one second. Another significant feature to note is that sensitivity starts dropping at negative adapting intervals, i.e., when the test flash precedes the onset of the adapting stimulus. The results from the other *S* were equally as consistent as these, and similar, except that no decrease in sensitivity occurred until an adapting interval of −.2 sec. was reached. Note also that there appears to be no discontinuity of the function as it passes through zero time.

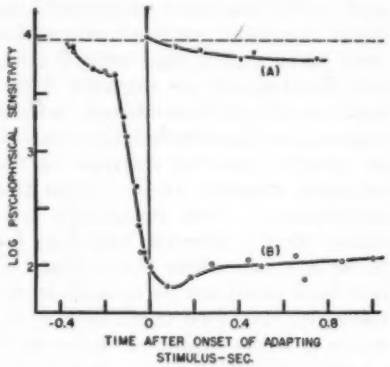


FIG. 3. Curve *A:* Sensitivity, determined .4 sec. after the cessation of the adapting stimulus, plotted as a function of the duration of pre-exposure. Curve *B:* Sensitivity changes as a function of adapting interval. The sensitivity values represent the density of neutral filters used. The negative adapting intervals mean that the test flash preceded the adapting stimulus.

Curve *A* in Fig. 3 shows sensitivity as determined .4 sec. after the cessation of the adapting stimulus, the latter having been in effect for the time indicated on the abscissa. This function shows that recovery was nearly complete after only .4 sec.

ERG determinations.—It is often desirable, when one is concerned with the comparison of psychophysical and ERG data, to utilize the concept of a *criterion response*, so that a "retinal threshold" may be obtained. In the present work, a peculiar problem arises because the response to the test flash does not always appear as a discrete event on the record. It would be desirable to determine the relative test-flash luminance required to elicit a criterion response at each adapting interval. However, adapting stimuli which are able measurably to reduce the height of the *b*-wave to a subsequent test flash must also be capable of eliciting measurable responses themselves.[6] Therefore, when we examine a typical record, we see not only the response to the test flash, but also a response to the adapting stimulus. If the test-flash response comes clearly before, or after, the response to the adapting stimulus, the former is easily measured and the stimulus conditions leading to a criterion response may be determined. On the other hand, if the onsets of the test flash and test stimuli are within .15 sec. (in either direction), the two responses interact, and a direct measure of the response height to the test flash becomes impossible. This phenomenon is illustrated

been used, it would have been (and was, in some exploratory experiments) possible to obtain electrical data at luminance levels closer to those employed in the psychophysical experiment. However, such data would be encumbered by lack of control over pupillary aperture, since artificial pupils are not appropriate for extremely large stimulus fields. With the present arrangement, it is absolutely certain that the same receptors were stimulated by the adapting stimulus as by the test flash, and pupillary changes have no effect.

[6] Space limitations make it impossible to include full data in support of this generalization, which was found to hold over a much wider range of stimulus conditions than those involved here.

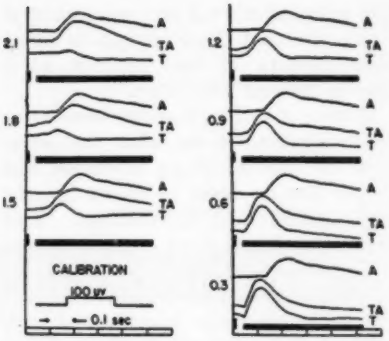


FIG. 4. Tracings of original records, showing the response interaction to stimuli which do not physically overlap in time. In this case, the test flash preceded the adapting stimulus onset by .04 sec. (an "adapting interval" of −.04 sec.). The numbers to the left of each set of responses represent the number of log units of neutral density filtering used in the test-flash system. Two log units of neutral filtering were used at all times in the adapting system. The letters to the right of the responses indicate the following: *A*—response to adapting stimulus alone; *TA*—response to combined adapting stimulus and test stimulus (whose durations are shown below each set of records); *T*—response to test flash alone.

in Fig. 4. In this example, the test flash preceded the onset of the adapting stimulus by .04 sec. At the lowest test-flash luminance, the response to the combined test flash–adapting stimulus sequence differs little from that to the adapting stimulus alone. At the highest test-flash luminance, the response to the combination is almost identical to that produced by the test flash alone. Somewhere within these limits, we would like to select an intensity value which represents the retinal threshold.

This was accomplished with the help of an auxiliary experiment, where *S*'s report and ERG data were obtained simultaneously. In order to make a meaningful comparison, a very large area of stimulation had to be used. Such an area was obtained by removing lens L_7 (Fig. 1) and allowing the collimated beam to enter an integrating sphere, which also had an aperture through which *S* could look.[7] The adapting luminance was adjusted so that it elicited a response of magnitude equivalent to those obtained in the main experiment. Responses were obtained at various test-flash luminances, for an adapting interval of .2 sec. At this interval, responses to the test flash were clearly visible following the response to adapting stimulus onset. It was found, under this condition, that *S* could not discriminate an intensity increment until it produced a *b*-wave response of 30 μv in height. Next, adapting intervals of +.105 sec. and −.042 sec. were employed. At these intervals, response interaction took place, and in each case, *S* could not discriminate an intensity increment until the combined *b*-wave response to the combination was 60 μv greater than the response to the adapting stimulus alone. From this experiment, it was tentatively concluded that a criterion response, for those intervals where the response to test flash could not be independently measured, could be taken as $x + 2h$, where x is the height of response to adapting stimulus alone, and h the criterion response height selected when the test-flash response could be isolated and directly measured.

On this basis, the points shown in Fig. 5 were derived. These are data from one *S;* the other *S* yielded very similar results. The similarity between the psychophysical data of Fig. 3 and the ERG data of Fig. 5 strongly suggests a retinal basis for part of the psychophysical results.

Longer time intervals.—In one of the

[7] The authors wish to thank Dr. E. Parker Johnson, who built the integrating sphere and also served as one of the *S*s. The short adapting intervals used in this auxiliary experiment negate the importance of pupillary effects.

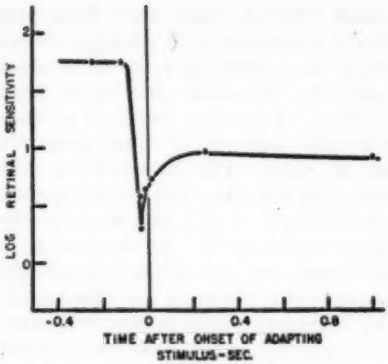


FIG. 5. Retinal sensitivity as a function of adapting interval

psychophysical sessions, *S* continued to receive the adapting stimulus for a period of 10 min., in order to determine what changes in the intensity-discrimination threshold occurred after the first second. After 5 min., sensitivity had *risen* .4 log units (compared to the 1-sec. value) and after 10 min., had dropped .05 log units from the 5-min. value. These exact figures are, of course, dubious, since they represent single determinations; however, there is almost no doubt that they agree with Baker's results on a qualitative level.

Curve *A* in Fig. 6 shows *b*-wave height as a function of time for a period of 10 min., as determined in another auxiliary experiment. Note that the initial drop, which takes place before zero time, is enormous, that there is a rise for 1 min., then a slow decline ensues to what looks like a steady state. Since increase in *b*-wave height to a constant test flash represents an increase in sensitivity, the data are again in accord with Baker's results.

ERG recovery data.—Figure 6 also

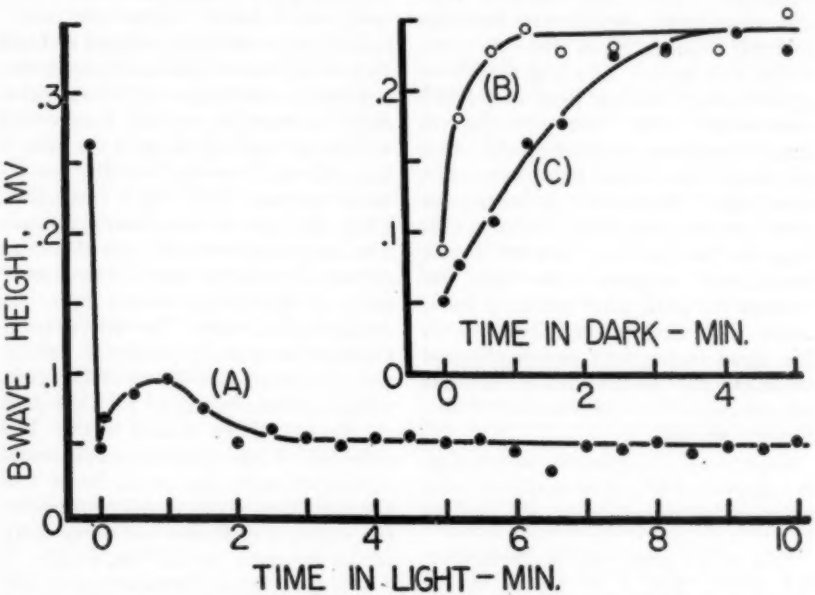


FIG. 6. *B*-wave height to a constant test flash as a function of (*A*) time in the light, (*B*) time in the dark, following 45 sec. in the light, and (*C*) time in the dark, following 10 min. in the light

shows recovery functions for two times of pre-exposure. Curve *B* shows recovery (*b*-wave height to a constant test flash as a function of time in the dark following adaptation) after 45-sec. exposure. Curve *C* shows recovery after 10-min. exposure. The differences between the curves are great, and are what would be expected on a photochemical basis. Nevertheless, note that *b*-wave height, determined while the adapting stimulus is still present (Curve *A*) changes relatively little over this period.

Discussion

The results of the present experiment are clearly in accord with the data of Baker (1) and Crawford (5). Our psychophysical data have shown the "Crawford effect," where a significant decrease in sensitivity occurs when a test flash precedes an adapting stimulus.[8] The data suggest that this effect begins earlier (.02 sec.) for scotopic vision, than for photopic vision (.01 sec.). This is generally in accord with reaction time and ERG data which show that photopic response processes are more rapid. Our psychophysical data have also indicated that sensitivity, following an initial tremendous decline, rises, rather than declines further. The ERG data tentatively suggest that this rise reaches its peak after about 1 min., under the restricted conditions used. We have incomplete psychophysical evidence also on this point, but the data obtained are quite in accord with Baker's results.

At normal stimulating intensities, the human ERG is a response only to stimulus onset. Our ERG data show that a weak test flash produces a response with a longer latency than that resulting from a stronger adapting stimulus, and that both response processes occupy a considerable amount of time (from .1 to .2 sec.). The response to the adapting stimulus completely engulfs the response to the test flash, if the latter is too small. If the test flash is intense enough, it correspondingly engulfs the response to the adapting stimulus. This interaction effect occurs at adapting intervals identical to those where the "Crawford effect" was found in the psychophysical experiment.

A study by Riggs and Graham (9) provides significant evidence on light adaptation at the neural level. These investigators, using single nerve-fiber recording techniques, found that the onset of illumination is followed by an initially high rate of activity in the optic nerve fibers. They also established that sensitivity, defined as $1/\Delta I$ (where ΔI was the intensity increment required to produce a criterion increment increase in response frequency) was low at the beginning of the adapting process, increased rapidly for 20 to 80 sec., and then began to decline. They did not use very short or negative adapting intervals, and therefore did not investigate the phenomenon of primary sensitivity decline from the dark-adapted state. The order of magnitude of sensitivity increase which they observed was about .4 log unit, which is nicely in accord with the data of Baker. They stated that "The possibility suggests itself that processes associated with the initial burst and the silent period may mask or suppress the response to flashes occurring early in the exposure period" (9, p. 22).

Our theoretical position is generally in accord with this suggestion given by Riggs and Graham, which is similar

[8] Not to be confused with the "Stiles-Crawford effect," which is an entirely different phenomenon. We have coined the term "Crawford effect" merely to serve as a shorthand designation.

to that of Baker and Crawford given at the outset of this paper. The adaptation process with which we have been concerned might best be considered simply as some kind of nonphotochemical adaptation. In the most general terms, the fibers of the optic nerve may be considered to fire at such a high frequency to the onset of the adapting stimulus that a stimulus capable of providing a discriminable frequency increment after a minute or so of adaptation (when the frequency of response to the adapting stimulus is lower) is not able to supply such an increment, although a higher intensity stimulus can. Until we have more complete information on the relationship between stimulus intensity, photoreceptor activity, and frequency of nerve-fiber discharge (and, in particular, what is a discriminable frequency increment) there is little to be gained by further speculation.

Neither the "Crawford effect," nor Baker's secondary-sensitivity rise has ever been observed in experiments on light adaptation involving binocular matching. This is what one would expect on the basis of a nonphotochemical hypothesis, since the eye receiving the test flash, in such binocular experiments, is in the dark, or in any event does not receive and respond to the adapting stimulus. It is perhaps important to note here that Schouten and Ornstein's (**10**) so-called "alpha adaptation" was discovered with a binocular technique, and therefore probably represents a phenomenon different from the rapid initial sensitivity drop which we have noted here.

It is perhaps of special significance that we have been able to duplicate the results of Baker and Crawford with such extremely weak adapting stimuli. It is almost inconceivable that enough photosensitive substance could be consumed, at our adapting level and short adapting intervals, to account for the initial sensitivity drop observed.[9] What little recovery data we obtained supports this contention.

If this analysis is correct, the present study has provided us with a method whereby neural activity, at a peripheral level, may be indirectly measured, using psychophysical methods and human *Ss*. In addition to extending this method to new problems, it is hoped that a more rigorous theoretical conception may follow from further ERG-psychophysical comparisons obtained at several adapting levels, longer adapting intervals, and various adapting areas.

Summary

Two adaptation phenomena, previously isolated at photopic adapting levels by psychophysical methods, have been shown to hold at scotopic adapting intensities. These are: (*a*) The "Crawford effect," where the threshold to a test flash is increased when it *precedes* an adapting stimulus. It is shown that if such "backward" adapting intervals are not used, an initial, and relatively enormous, drop in sensitivity following adapting stimulus onset will not show in the data. (*b*) A secondary rise in sensitivity following the initial decrease, which is small compared to the initial drop, and was first shown by Baker.

[9] Exploratory research indicates that the magnitude of the initial sensitivity decline would have been considerably reduced, had the test flash subtended a smaller visual angle than the adapting stimulus. Identical areas were deliberately employed in this experiment, in order that the same receptors would be stimulated by each stimulus, thus enabling a direct comparison between the psychophysical and ERG data. In the sense that identical areas eliminate the cue of contour, identical areas perhaps provide for a "purer" intensity discrimination. The variable of test-flash area remains an interesting one for future exploration.

ERG data from human *Ss* are also given which strengthen the interpretations put upon the aforementioned phenomena by Baker and Crawford. It is concluded that the initial decrease in sensitivity has a nonphotochemical basis, perhaps in the initially rapid response of the optic nerve fibers to adapting stimulus onset, and that the secondary rise reflects the slowing of such activity. The "Crawford effect" at scotopic levels appears in the ERG data as an interaction between response processes (*b*-waves) which clearly overlap in time, in spite of the temporal gap between the stimuli which produced them.

References

1. Baker, H. de H. The course of foveal light adaptation as measured by the intensity threshold increment. *J. opt. Soc. Amer.*, 1949, **39**, 172–179.
2. Baker, H. de H. A suggested analysis of the intensity discrimination light adaptation process. *Amer. Psychologist*, 1951, **6**, 278–279.
3. Boynton, R. M. Stray light and the human electroretinogram. *J. opt. Soc. Amer.*, in press.
4. Boynton, R. M., & Riggs, L. A. The effect of stimulus area and intensity upon the human retinal response. *J. exp. Psychol.*, 1951, **42**, 217–226.
5. Crawford, B. H. Visual adaptation to brief conditioning stimuli. *Proc. Roy. Soc. B.*, 1947, **134**, 283–302.
6. Hartline, H. K., & McDonald, P. R. Light and dark adaptation of single photoreceptor elements in the eye of *Limulus*. *J. cell. and comp. Physiol.*, 1947, **30**. 225–254.
7. Riggs, L. A. Continuous and reproducible records of the electrical activity of the human retina. *Proc. Soc. exp. Biol. Med.*, 1941, **48**, 204–207.
8. Riggs, L. A., Berry, R. N., & Wayner, M. A comparison of electrical and psychophysical determinations of the spectral sensitivity of the human eye. *J. opt. Soc. Amer.*, 1949, **39**, 427–436.
9. Riggs, L. A., & Graham, C. H. Some aspects of light adaptation in a single photoreceptor unit. *J. cell. comp. Physiol.*, 1940, **16**, 15–23.
10. Schouten, J. F., & Ornstein, L. S. Measurements of direct and indirect adaptation by means of binocular vision. *J. opt. Soc. Amer.*, 1939, **29**, 168–192.
11. Wald, G., & Clark, A. B. Visual adaptation and the chemistry of the rods. *J. gen. Physiol.*, 1937, **21**, 93–105.

(Received November 24, 1952)

Journal of Experimental Psychology
Vol. 46, No. 2, 1953

COMMENT ON "THE ROLE OF INSTRUCTION IN EXPERIMENTAL SPACE ORIENTATION"

H. A. WITKIN

College of Medicine, State University of New York

In a recent issue of this Journal (1) Mann and Boring reported a study that appears to have been based on a serious misunderstanding of the procedures used in our laboratory to investigate problems of space orientation.

The Mann-Boring study was concerned with the effect of different kinds of instructions upon *S*'s performance in establishing the upright in a rod-and-frame type of situation. Two groups of four *S*s each were used, a naive group and a sophisticated group. The *S* in the naive group was simply told that he was to set the target rod to the vertical; "no instructions were given with respect to what constituted verticality, and if he asked questions, he was told to use his own judgment in setting the target rod" (1, p. 45). The *S* in the sophisticated group was specifically told to set the target "to the gravitational vertical, that is, either parallel to the walls of the room, or at right angles to the floor" (1, p. 45), and he was also given a few practice trials "in order to make sure that he was aware of the criterion of verticality" (1, pp. 45–46). The results showed that the mean deviation of the rod from the true upright and group variability (in a group of four *S*s) was greater for the naive group than for the sophisticated group.

The Mann-Boring report as a whole clearly indicates the authors' assumption that the instructions given their naive group corresponded more or less to those we have been using in our laboratory, whereas the instructions given their sophisticated group represented those employed in the studies of Mann and his co-workers at Tulane. Moreover, Mann and Boring suggest that the differences in results between the Tulane studies and our own are attributable to differences in instruction used, and that the individual differences reported from our laboratory "are probably a function of the level of instructions given *S*" (1, p. 47).

The impression offered by Mann and Boring of the instructions used in our experiments is not a correct one. The instructions we have been using bear essentially no resemblance to those given by Mann and Boring to their naive group, and in fact they correspond to those given the sophisticated group. To illustrate our procedure, we may cite the following verbatim account of instructions to *S* from our most recent report (2) which, together with a number of other reports, was available to Mann and Boring at the time their manuscript was submitted. These instructions are representative of those we have used throughout.

> Let me tell you at this point what I mean by straight. When your chair is straight, it should be lined up with the walls of Ingersoll [the laboratory building] outside, and you should be sitting straight up and down the way you are right now. When the room is straight, its walls should be lined up with the walls of Ingersoll outside, the way they are now, and its floor should be straight with the floor of Ingersoll outside, the way it is now. In other words, when the room is straight, it should be standing exactly the way it is right now. Are there any questions?

Furthermore, after the first trial, which was a chair-adjustment trial, *S* was asked, "Is your chair now straight with the walls of the building out here,

so that you are sitting straight up and down the way you were when you came in?" This question was repeated after every single chair-adjustment trial. After every room-adjustment trial, *S* was asked, "Are the walls of the room now straight with the walls of the building out here, and is the floor of the room that you stood on when you came in now straight with the floor out here, so that the room is standing the way it was when you came in?"

Using this kind of explicit definition of the upright, characterized always as corresponding to the gravitational upright, we have observed marked individual differences in performance in all the situations studied. These correspond to real differences in perception, and in this respect are probably different from the individual differences found by Mann and Boring in their naive group, which very likely reflect differences in the decisions made by *S*s—left uninstructed with regard to their real task—as to what *E* expected of them.

The methodological issue involved in the problem of instructions to *S* is, of course, an important one, and there is no doubt that the kind of instructions used may influence performance in perceptual situations and may affect the nature and extent of individual differences obtained. At the same time, it is clear that the differences between Mann's results and ours (with our studies, in general, providing evidence of greater variation among individuals in mode of space orientation and greater influence by the visual field) cannot be attributed to differences in instructions used. Nor can the phenomenon of individual differences in perception be dismissed by attributing these differences to inadequate instruction of *S*s. Further evidence of the importance and deep-seated nature of individual differences in manner of perceiving will be set forth in the next report from our laboratory, concerned with stability in mode of perception over extended periods of time, self-consistency in perception under varied conditions, and the relation between perception and personality organization.

References

1. Mann, C. W., & Boring, R. O. The role of instruction in experimental space orientation. *J. exp. Psychol.*, 1953, **45**, 44–48.
2. Witkin, H. A. Further studies of perception of the upright when the direction of the force acting on the body is changed. *J. exp. Psychol.*, 1952, **43**, 9–20.

(Received May 13, 1953)

PSYCHOLOGICAL MONOGRAPH

#270

1945

ON PROBLEM SOLVING

By

KARL DUNCKER

$2.50

This popular monograph is #270 of the Psychological Monograph series. It has been reprinted so that it is again available.

Third Printing

AMERICAN PSYCHOLOGICAL ASSOCIATION

1333 Sixteenth Street N. W.

Washington 6, D. C.

PSYCHOLOGICAL MONOGRAPHS: GENERAL AND APPLIED

Volume 66, 1952

STUDIES IN REPUTATION: I. SEX AND GRADE DIFFERENCES IN SCHOOL CHILDREN'S EVALUATIONS OF THEIR PEERS. II. THE DIAGNOSIS OF SOCIAL ADJUSTMENT. Read D. Tuddenham. #333. $1.50.

RELIABILITY AND VALIDITY OF JUDGES' RATINGS OF ADJUSTMENT ON THE RORSCHACH. Marguerite Q. Grant, Virginia Ives, and Jane H. Ranzoni. #334. $1.00.

SOME PERSONALITY CORRELATES OF RELIGIOUS ATTITUDES, AS DETERMINED BY PROJECTIVE TECHNIQUES. Ralph Mason Dreger. #335. $1.00.

PERSONAL SECURITY AS RELATED TO STATION IN LIFE. Andie L. Knutson. #336. $1.00.

THE VALIDITY OF PERSONALITY-TRAIT RATINGS BASED ON PROJECTIVE TECHNIQUES. Henry Samuels. #337. $1.00.

RORSCHACH ESTIMATES OF PERSONALITY ATTRIBUTES IN THE MICHIGAN ASSESSMENT PROJECT. Woodrow Wilbert Morris. #338. $1.00.

DEVELOPMENT OF AN EYE CAMERA FOR USE WITH MOTION PICTURES. Paul R. Wendt. #339. $1.00.

FACTOR ANALYSIS OF SOME REASONING TESTS. Harold M. Corter. #340. $1.00.

THE USE OF VOCATIONAL INTEREST SCALES IN PLANNING A MEDICAL CAREER. Edward K. Strong, Jr., and Anthony C. Tucker. #341. $2.00.

THE OCCUPATIONAL LEVEL SCALE AS A MEASURE OF DRIVE. Gordon J. Barnett, Irving Handelsman, Lawrence H. Stewart, and Donald E. Super. #342. $1.50.

OBJECTIVE MEASUREMENT OF REALITY-CONTACT WEAKNESS. Albert Kreinheder. #343. $1.00.

THE PERFORMANCE OF SCHIZOPHRENIC AND NORMAL INDIVIDUALS FOLLOWING FRUSTRATION. Harold Wilensky. #344. $1.00.

PARENTS' ATTITUDES VS. ADOLESCENT HOSTILITY IN THE DETERMINATION OF ADOLESCENT SOCIOPOLITICAL ATTITUDES. Kenneth Helfant. #345. $1.00.

PERSONALITY OF THE CONVULSIVE PATIENT IN MILITARY SERVICE. T. W. Richards. #346. $1.00.

THE ACADEMIC ACHIEVEMENT OF VETERAN AND NONVETERAN STUDENTS. Norman Frederiksen and W. B. Schrader. #347. $1.50.

THE PRIMARY PERSONAL VALUES MEASURED BY THE ALLPORT-VERNON TEST, "A STUDY OF VALUES." Hubert E. Brogden. #348. $1.00.

SELECTED PERSONALITY VARIABLES AND THE LEARNING PROCESS. Eugene L. Gaier. #349. $1.00.

JUDGING INTERESTS FROM EXPRESSIVE BEHAVIOR. N. L. Gage. #350. $1.00.

Orders for any of these Monographs can be placed separately at the prices listed above, or the entire volume can be ordered for $6.00.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1333 Sixteenth Street N. W.

Washington 6, D. C.

ED

OOL
OSIS

ON
H.

DE-
ger.

nut-

TIVE

CHI-
1.00.
RES.

340.

ICAL
2.00.
rdon
uper.

lbert

FOL-

INA-
Hel-

VICE.

STU-

RNON
00.
CESS.

350.

listed